GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 03:15:53 ; Search time 8364 Seconds
(without alignments)
11136.320 Million cell updates/sec

Title: US-10-027-880-1
Perfect score: 2149
Sequence: 1 gtcgaattgtgatatattgt.....actaattgctaattaatcag 2149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

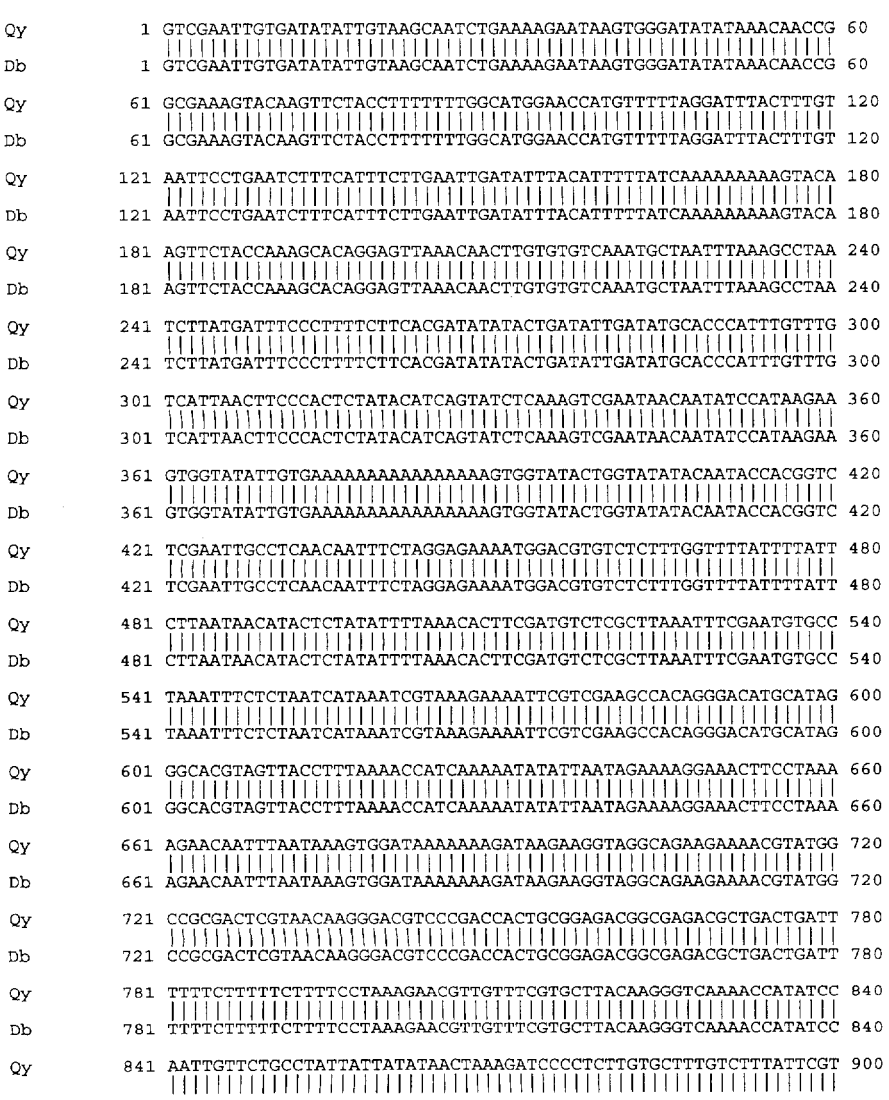
Result No.		Score	% Query Match	Length	DB	ID	Description
1		2149	100.0	2149	6	AX063409	AX063409 Sequence
2		2149	100.0	4280	6	AX063410	AX063410 Sequence
3		2149	100.0	4309	6	AX063412	AX063412 Sequence
4		2149	100.0	4413	6	AX063411	AX063411 Sequence
5		2149	100.0	4413	8	AF242314	AF242314 Arabidops
c 6		2149	100.0	82454	8	AC012193	AC012193 Arabidops
7		1733	80.6	1741	8	AF242313	AF242313 Arabidops
c 8		251.8	11.7	317	8	ATH553603	AJ553603 Arabidops
9		132	6.1	295	11	AL773135	AL773135 Arabidops
c 10		93.8	4.4	43982	8	AC007289	AC007289 Arabidops
11		76.2	3.5	343050	3	PFA929353	AL929353 Plasmodiu
c 12		71	3.3	175842	9	AL731547	AL731547 Human DNA
13		70.4	3.3	14867	3	AE001398	AE001398 Plasmodiu
14		70.4	3.3	347050	3	PFA929351	AL929351 Plasmodiu
c 15		69.2	3.2	104386	8	ATT32A17	AL161813 Arabidops
16		69.2	3.2	170627	2	AC125567	AC125567 Rattus no
c 17		69.2	3.2	179771	8	ATCHRIV25	AL161513 Arabidops
c 18		69	3.2	175544	2	AC117342	AC117342 Rattus no
19		68.4	3.2	250707	3	AE014848	AE014848 Plasmodiu
20		66.8	3.1	129240	9	AC084128	AC084128 Homo sapi
c 21		66.2	3.1	70098	8	AB025605	AB025605 Arabidops
c 22		66.2	3.1	73663	8	AP002029	AP002029 Arabidops
23		66	3.1	181792	9	AC098822	AC098822 Homo sapi
24		65.6	3.1	154995	9	AC011979	AC011979 Homo sapi
25		65.2	3.0	249995	3	AE014840	AE014840 Plasmodiu
c 26		64.6	3.0	111861	9	AC069435	AC069435 Homo sapi
27		64.6	3.0	156060	2	AC004153	AC004153 Plasmodiu
28		64.4	3.0	20529	2	AC116922	AC116922 Dictyoste
29		64.2	3.0	152824	2	BX663616	BX663616 Danio rer
30		64	3.0	144076	5	AL928677	AL928677 Zebrafish
31		63.8	3.0	183584	9	AC012492	AC012492 Homo sapi
c 32		63.8	3.0	189402	9	AC083829	AC083829 Homo sapi
c 33		63.6	3.0	7008	6	AX251286	AX251286 Sequence
34		63.4	3.0	526	9	HSFRAN1	AF053528 Homo sapi
35		63.4	3.0	110000	2	PFMAL7P1_07	Continuation (8 of
36		63.4	3.0	160661	2	BX548246	BX548246 Danio rer
c 37		63.2	2.9	148337	2	AC021509	AC021509 Homo sapi
c 38		63.2	2.9	171255	9	AL691447	AL691447 Human DNA
39		62.8	2.9	125623	3	AC115599	AC115599 Dictyoste
c 40		62.6	2.9	331039	3	AC116988	AC116988 Dictyoste
41		62.4	2.9	110000	2	PFMAL13_00	AL844509 Plasmodiu
c 42		62.4	2.9	110000	2	PFMAL13_23	Continuation (24 o
c 43		62.4	2.9	154890	2	BX548249	BX548249 Danio rer
44		62.2	2.9	1141	6	AX083744	AX083744 Sequence
c 45		61.8	2.9	139111	9	AL158048	AL158048 Human DNA

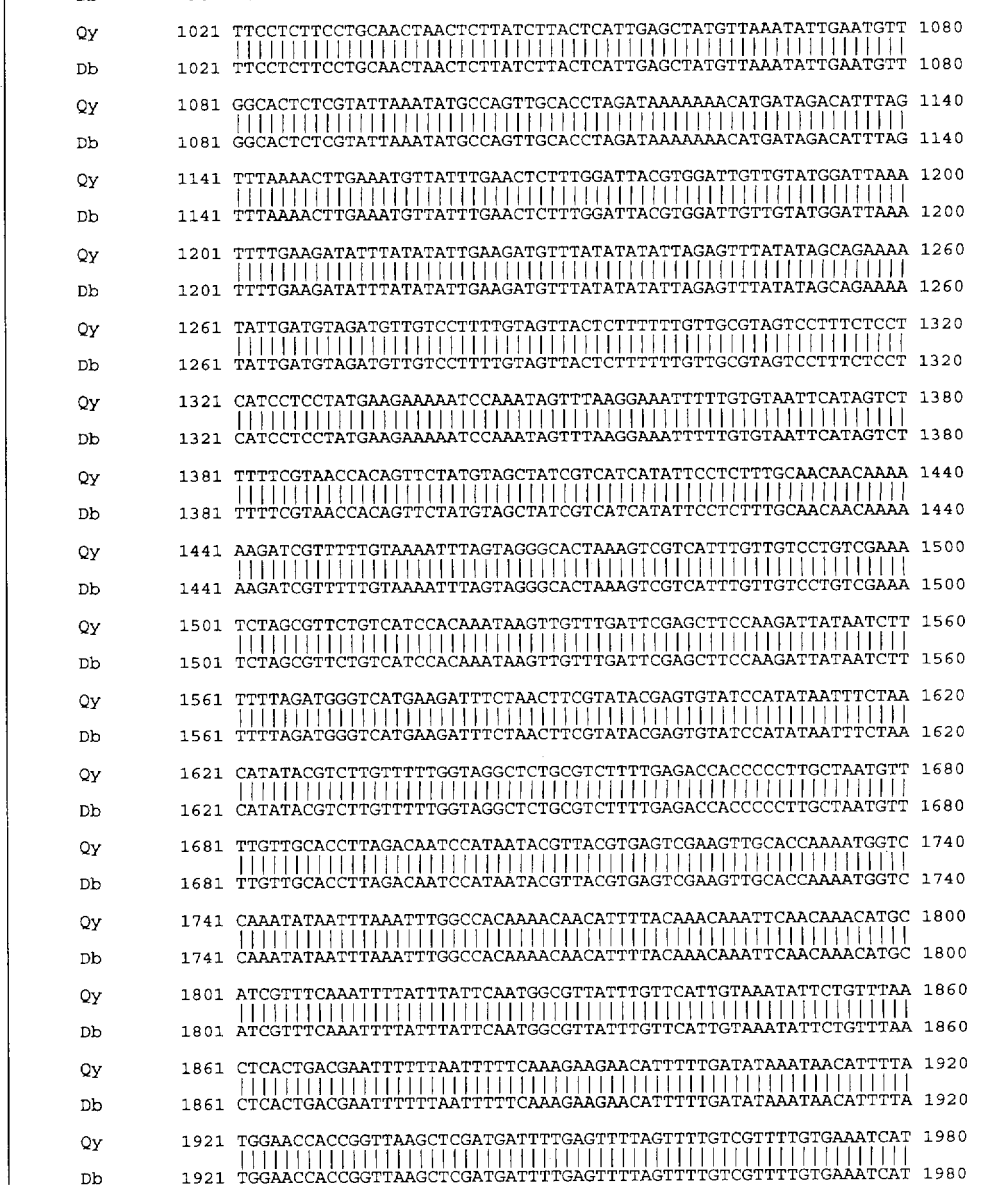
ALIGNMENTS

RESULT 1
AX063409
LOCUS AX063409 2149 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 1 from Patent WO0100833.
ACCESSION AX063409
VERSION AX063409.1 GI:12541197
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Hoffmann,B., Mollier,P. and Pelletier,G.
TITLE Promoter expressed specifically in the cells of plant roots,

recombinant vectors and host cells comprising same and transgenic plants obtained
JOURNAL Patent: WO 0100833-A 1 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES Location/Qualifiers
source 1. .2149
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
ORIGIN

Query Match 100.0%; Score 2149; DB 6; Length 2149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





```
Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149


RESULT 2
AX063410
LOCUS       AX063410               4280 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 2 from Patent WO0100833.
ACCESSION   AX063410
VERSION     AX063410.1  GI:12541198
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Hoffmann,B., Mollier,P. and Pelletier,G.
  TITLE     Promoter expressed specifically in the cells of plant roots,
            recombinant vectors and host cells comprising same and transgenic
            plants obtained
  JOURNAL   Patent: WO 0100833-A 2 04-JAN-2001;
            INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES             Location/Qualifiers
     source          1. .4280
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Construction promoteur + sequence codante du g ne
                     gus"
ORIGIN

  Query Match             100.0%;  Score 2149;  DB 6;  Length 4280;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAACATGATAGACATTTAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAACATGATAGACATTTAG 1140

Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149


RESULT 3
AX063412
LOCUS       AX063412               4309 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 4 from Patent WO0100833.
ACCESSION   AX063412
VERSION     AX063412.1  GI:12541200
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Hoffmann,B., Mollier,P. and Pelletier,G.
  TITLE     Promoter expressed specifically in the cells of plant roots,
            recombinant vectors and host cells comprising same and transgenic
            plants obtained
  JOURNAL   Patent: WO 0100833-A 4 04-JAN-2001;
            INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES             Location/Qualifiers
     source          1. .4309
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Insert du vecteur pBin19"
ORIGIN

  Query Match             100.0%;  Score 2149;  DB 6;  Length 4309;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         30 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 89

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 149

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 209

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 269

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 329

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 389

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 449

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 509

Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 569

Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 629

Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 689

Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 749

Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 809

Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 869

Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 929

Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 989

Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1049

Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1109

Qy       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

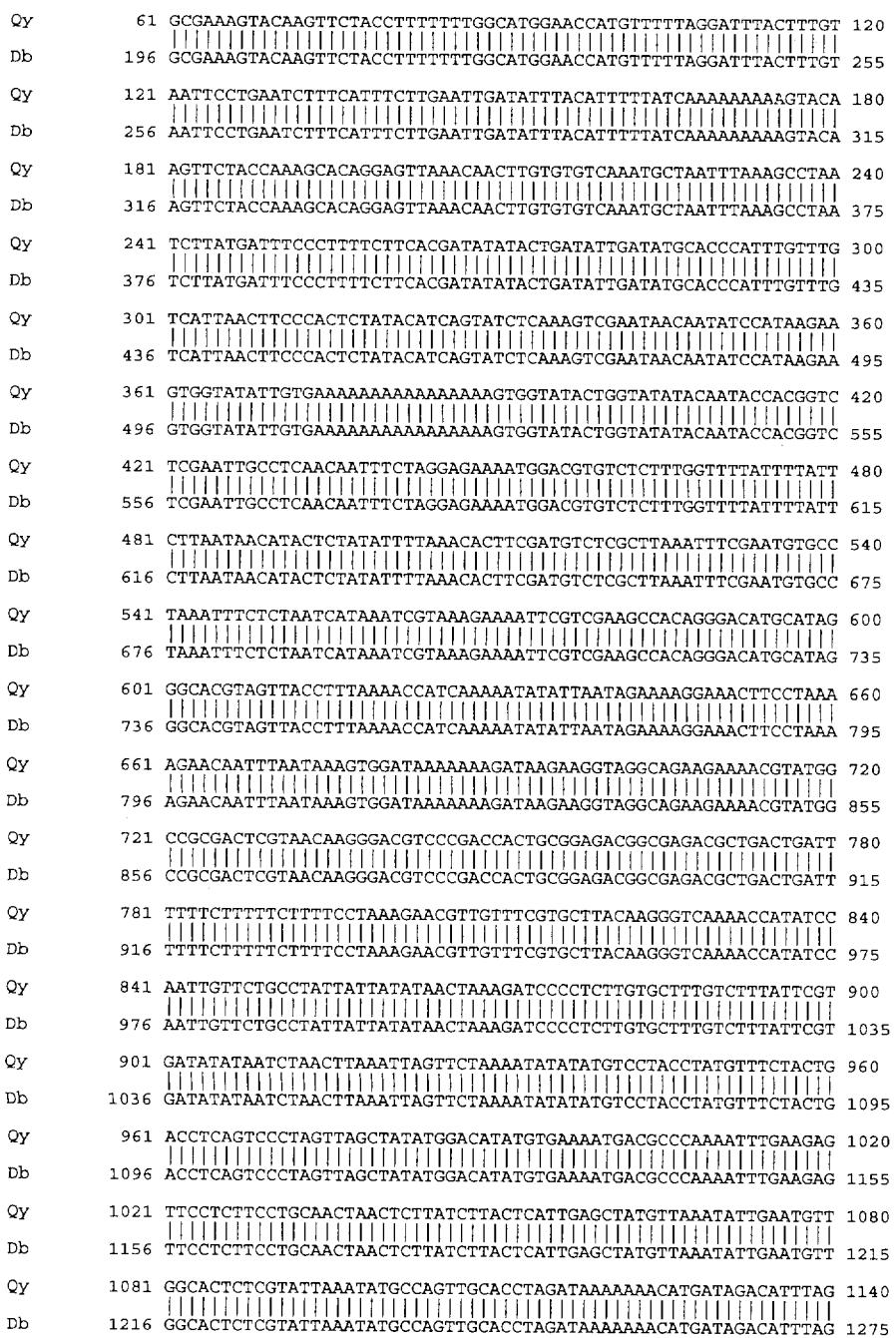
```
Db       1110 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1169
Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
Db       1170 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1229
Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
Db       1230 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1289
Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
Db       1290 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1349
Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
Db       1350 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1409
Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
Db       1410 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1469
Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
Db       1470 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1529
Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
Db       1530 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1589
Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
Db       1590 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1649
Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
Db       1650 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1709
Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
Db       1710 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1769
Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
Db       1770 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1829
Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
Db       1830 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1889
Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
Db       1890 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1949
Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
Db       1950 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2009
Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
Db       2010 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2069
Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
Db       2070 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2129
Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
Db       2130 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2178
```

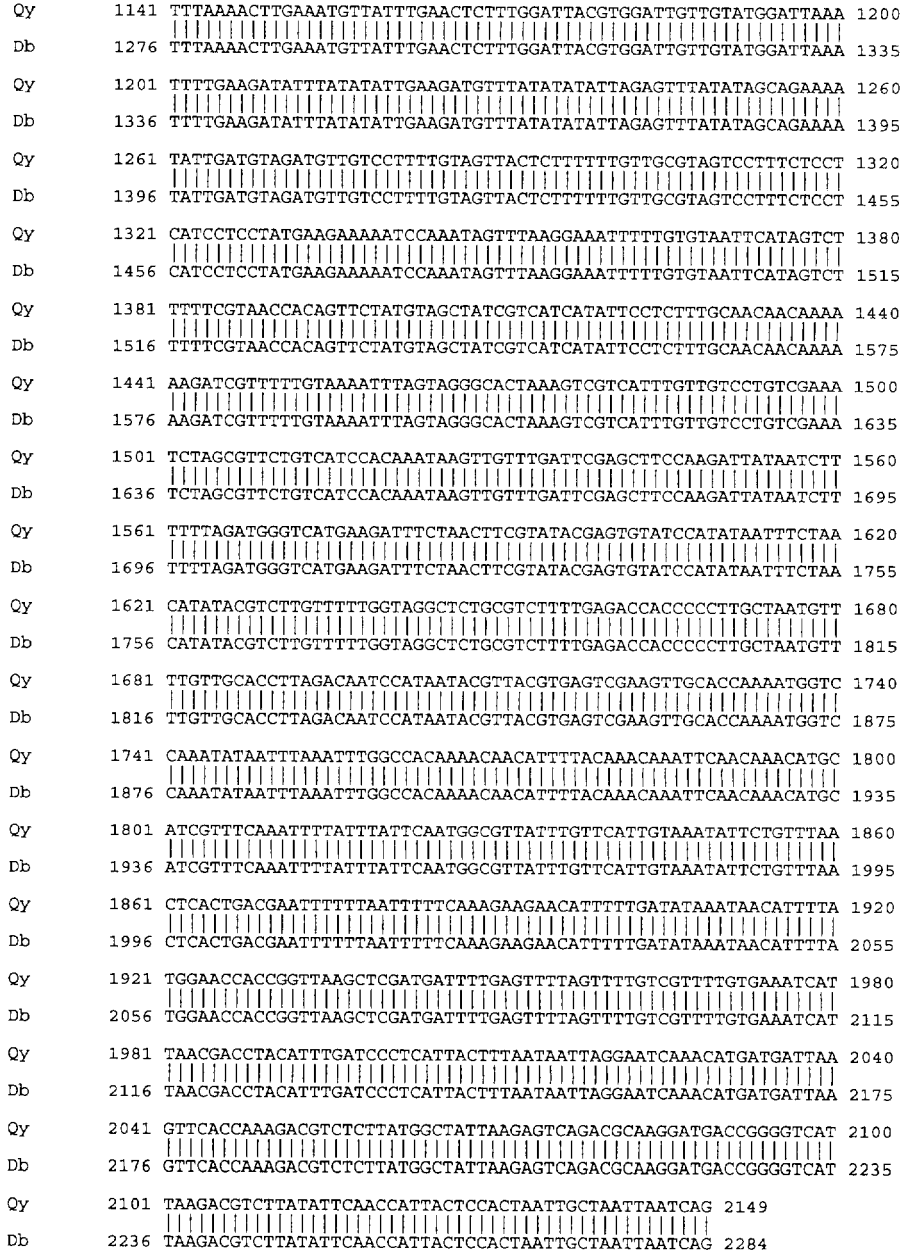
RESULT 4
AX063411

```
LOCUS       AX063411                4413 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 3 from Patent WO0100833.
ACCESSION   AX063411
VERSION     AX063411.1  GI:12541199
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Hoffmann,B., Mollier,P. and Pelletier,G.
  TITLE     Promoter expressed specifically in the cells of plant roots,
            recombinant vectors and host cells comprising same and transgenic
            plants obtained
  JOURNAL   Patent: WO 0100833-A 3 04-JAN-2001;
            INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES             Location/Qualifiers
     source          1. .4413
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

  Query Match           100.0%;  Score 2149;  DB 6;  Length 4413;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
Db        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195
Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
Db        196 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255
Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
Db        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315
Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
Db        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375
Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
Db        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435
Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
Db        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
Db        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555
Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
Db        556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615
Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
Db        616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675
Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
Db        676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735
Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
Db        736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795
Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
```

```
Db        796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855
Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
Db        856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915
Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
Db        916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975
Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
Db        976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035
Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
Db       1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095
Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
Db       1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155
Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
Db       1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215
Qy       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
Db       1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275
Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
Db       1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335
Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
Db       1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395
Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
Db       1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455
Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
Db       1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515
Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
Db       1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575
Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
Db       1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635
Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
Db       1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695
Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
Db       1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
Db       1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815
Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
Db       1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875
Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
Db       1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935
Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
Db       1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995
Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
Db       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
Db       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115
Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
Db       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175
Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
Db       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235
Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
Db       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
```

```
RESULT 5
AF242314
LOCUS       AF242314                4413 bp    DNA     linear   PLN 06-AUG-2000
DEFINITION  Arabidopsis thaliana ribosomal S13-like protein gene, partial cds;
            nuclear gene for chloroplast product.
ACCESSION   AF242314
VERSION     AF242314.1  GI:9719365
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4413)
  AUTHORS   Mollier,P., Hoffmann,B., Pelletier,G. and Chwetzoff,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Genetique, INRA, Route de St Cyr,
            Versailles 78026, France
FEATURES             Location/Qualifiers
     source          1. .4413
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="near m532"
                     /clone="BAC T32E8"
                     /ecotype="Columbia"
     mRNA            <3738. .>3806
                     /product="ribosomal S13-like protein"
     CDS             3738. .>3806
                     /note="similar to Arabidopsis thaliana ribosomal protein
                     S13 encoded by GenBank Accession Number Z47986"
                     /codon_start=1
                     /product="ribosomal S13-like protein"
                     /protein_id="AAF97767.1"
                     /db_xref="GI:9719366"
                     /translation="MLGLRRSATTLFDISQSLLRNVT"
ORIGIN

  Query Match           100.0%;  Score 2149;  DB 8;  Length 4413;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
Db        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195
```





RESULT 6
AC012193/c
LOCUS AC012193 82454 bp DNA linear PLN 19-JAN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence,

```
            complete sequence.
ACCESSION   AC012193
VERSION     AC012193.6  GI:12323286
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 82454)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82454)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 82454)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280880.
            Address all correspondence to:at@tigr.org

            BAC clone T32E8 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .82454
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="T32E8"
     gene            complement(1. .3320)
                     /gene="T32E8.1"
                     /note="similar to ribonuclease II RNB family protein
                     GB:CAB16367 [Schizosaccharomyces pombe]; Pfam HMM hit:
                     VacB and RNase II family 3'-5' exoribonucleases"
     mRNA            complement(join(<1. .69,154. .2280,2381. .2449,2524. .2577,
                     2676. .3055,3215. .>3320))
                     /gene="T32E8.1"
     CDS             complement(join(<1. .69,154. .2280,2381. .2449,2524. .2577,
                     2676. .3055,3215. .3320))
                     /gene="T32E8.1"
                     /codon_start=1
                     /product="putative 3'-5' exoribonuclease, 3' partial;
                     3320-1"
                     /protein_id="AAG51632.1"
                     /db_xref="GI:12323300"
                     /translation="MKSASSEQSVERIENGHKKKRNRPQKQNRRSKQSSVPIEDAHVE
                     ESLDGRDSSRSKAKDSTSSSKQQRPNTDELEAMRASNVAFNSMPPMRAESGYPRRSAS
                     PLLSSPEVSKQLLSKSCPDPRACEQSPGMNGELFQQIEGSSQRKIFSSHWSLDAVTEA
                     LEKGEAFKALFRVNAHNRNEAYCKIDGVPTDILINGNVCQSRAVEGDTVVIKLDPLSL
                     WPKMKGFVTESAAKPEGTNSPPEKDDKKARQKNGIDVVEGFEDGFSKNKSSVIGKGAK
                     NGVTPSSPPSLDSCLGSFCEQKGNCSAVDKLCGILSSFPHKRPTGQVVAVVEKSLVRD
                     SIVGLLDVKGWIHYKESDPKRCKSPLSLSDDEYVQLMPADPRFPKLIVPFHVLPGSIR
                     ARLENLDPNLEAELVAAQIVDWGEGSPFPVAQITHLFGRGSELEPQINAILYQNSVCD
                     SDFSPGSLTSLPRVPWEVPEEEVQRRKDLRDLCVLTIDPSTATDLDDALSVQSLPGGF
                     FRVGVHIADVSYFVLPETALDTEARFRSTSVYLMQRKISMLPPLLSENVGSLSPGADR
                     LAFSILWDLNREGDVIDRWIGRTIIRSCCKLSYDHAQDIIDGKSDVAENGWPALHGSF
                     KWCDVTRSVKQLSEISTTLRQKRFRNGALQLENSKPVFLFDEHGVPYDFVTCSRKGSN
                     FLVEEFMLLANMTAAEVISQAYRASSLLRRHPEPNTRKLKEFEGFCSKHGMDLDISSS
                     GQLQDSLEKITGNLKDDSVFVDILNNYAIKPMQLASYFCTGNLKDSVAEWGHYALAVP
                     LYTHFTSPLRRYPDIVVHRALAAALEAEELYSKQKQTAIDEGRSCFTGIHFNKDAAES
                     IEGKEALSVAALKHGVPSTEILSDVAAYCNERKLAARKVRDACDKLYTWFVLKQKEIF
                     PCEARVMNLGSRFMTVYISKL"
misc_feature         1. .189
                     /note="overlap with BAC clone T5M16
                     [AC010704:103165. .103353]."
gene                 complement(8589. .10716)
                     /gene="T32E8.2"
                     /note="similar to AUX1 (regulator of root gravitropism,
                     putative permease) GB:CAA67308 [Arabidopsis thaliana]"
mRNA                 complement(join(<8589. .8783,9000. .9465,9555. .9652,
                     9753. .9936,10020. .10132,10236. .10421,10504. .>10716))
                     /gene="T32E8.2"
CDS                  complement(join(8589. .8783,9000. .9465,9555. .9652,
                     9753. .9936,10020. .10132,10236. .10421,10504. .10674))
                     /gene="T32E8.2"
                     /codon_start=1
                     /product="putative AUX1-like permease; 10674-8589"
                     /protein_id="AAG51630.1"
                     /db_xref="GI:12323298"
                     /translation="MAAEKIETVVAGNYLEMEREEENISGNKKSSTKTKLSNFFWHGG
                     SVYDAWFSCASNQVAQVLLTLPYSFSQLGMMSGILFQLFYGLMGSWTAYLISVLYVEY
                     RTRKEREKFDFRNHVIQWFEVLDGLLGKHWRNLGLIFNCTFLLFGSVIQLIACASNIY
                     YINDKLDKRTWTYIFGACCATTVFIPSFHNYRIWSFLGLAMTTYTSWYLTIASLLHGQ
                     AEDVKHSGPTTMVLYFTGATNILYTFGGHAVTVEIMHAMWKPQKFKAIYLLATIYVLT
                     LTLPSASAVYWAFGDKLLTHSNALSLLPKTGFRDTAVILMLIHQFITFGFASTPLYFV
                     WEKLIGVHETKSMFKRAMARLPVVVPIWFLAIIFPFFGPINSAVGSLLVSFTVYIIPA
                     LAHMLTFAPAPSRENAVERPPRVVGGWMGTYCINIFVVVWVFVVGFGFGGWASMVNFV
                     RQIDTFGLFTKCYQCPPHKP"
gene                 12104. .13574
                     /gene="T32E8.3"
                     /note="similar to thaumatin-like protein GB:AAB71214
                     [Arabidopsis thaliana]; Pfam HMM hit: Thaumatin family"
mRNA                 join(<12104. .12179,12505. .12628,12704. .>13574)
                     /gene="T32E8.3"
CDS                  join(12104. .12179,12505. .12628,12704. .13574)
                     /gene="T32E8.3"
                     /codon_start=1
                     /product="thaumatin-like protein; 12104-13574"
                     /protein_id="AAG51631.1"
                     /db_xref="GI:12323299"
                     /translation="MDRTIEIEFDINGPYDSLIRPNISVASLTSRSENNITFDPKIFS
                     SKDANSFRFVTKDLVTLSHHIIRVWITIRTPVNYSTGTKWSESARVFTIINSCDQTIW
                     PAITPGENFNGGGFELKPGQSIVFHAPVGWSGRIWGRTGCKFDSTGTGTCETGSCGST
                     LKCSASGKPPASLAEFTLAALDFYDVSLVDGFNLPMSVTPMNGKGNCSVAGCVADLRP
                     HCPQELAVKSNGKVISCRSACDVFDRDEYCCRGVYGNPVVCQPTYYSKIFKQACPTAY
                     SYAYDDPTSIMTCTASDYVISFCSSRFVLFTFLTTSHDYFNYDSNQVSFILDHDFYIY
                     MIYFIHPLLISYAKVKVIQMQK"
gene                 14054. .15303
                     /gene="T32E8.4"
                     /note="similar to hypothetical protein GB:P34661
                     [Caenorhabditis elegans]"
mRNA                 join(14054. .14174,14251. .14348,14779. .14811,15170. .15303)
                     /gene="T32E8.4"
CDS                  join(14107. .14174,14251. .14348,14779. .14811,15170. .15252)
                     /gene="T32E8.4"
                     /codon_start=1
```

```
                     /product="unknown protein; 14107-15252"
                     /protein_id="AAG51633.1"
                     /db_xref="GI:12323301"
                     /translation="MATGGKVSFKVTLTSDPKLPFKVFSVPEGAPFTAVLKFAAEEFK
                     VPPQTSAIITNDGIGINPQQSAGNVFLKHGSELRLIPRDRVGAVFVMDP"
     gene            18583. .21388
                     /gene="T32E8.5"
                     /note="Pfam HMM hit: Eukaryotic protein kinase domain,
                     C-terminal region is identical to putative protein kinase
                     GB:CAA72680 [Arabidopsis thaliana]"
     mRNA            join(<18583. .18916,19056. .19948,20068. .20268,20382. .20543,
                     20662. .21388)
                     /gene="T32E8.5"
     CDS             join(18583. .18916,19056. .19948,20068. .20268,20382. .20543,
                     20662. .21234)
                     /gene="T32E8.5"
                     /codon_start=1
                     /product="putative protein kinase; 18583-21234"
                     /protein_id="AAG51619.1"
                     /db_xref="GI:12323287"
                     /translation="MQTTSIGPRRSVAPQRQASRNTRLVTAEGQRSQDVVTLSQSLAA
                     NTLTQDTTNNLAITSVAGESASITQPTVSEHFNPSDRQMDFGKSAVTSLESNLDVQRK
                     SQSLIGTSQDMEWDATNQAEASHLDACIGSKHQNLPSVDSEVSLKSEYKDSSSLAKIQ
                     GQLGEFPNFLNQPRTRCSAVGSSWATTTLIHSSSAPMLNATTHVSRSYVEADSNANPH
                     AVQSQGNLPSCCPSSKVSNILHPNKDATASEMPASTNDPEVRVKETDTSKQQQITTGL
                     EAPVGSSIYGSDGQANARLPEELHTSVSSQPQKSDKHEKVASSKGPSAPRKRNYDPDL
                     FFKVNGKLYQRLGKIGSGGSSEVHKVISSDCTIYALKKIKLKGRDYATAYGFCQEIGY
                     LKKLKGKTNIIQLIDYEVTDKTLLQEVLNGTMSNKDGRVKEDGFIYMVLEYGEIDLAH
                     MLSQKWREIEGSDRTIDENWLRFYWQQILQAVNTIHEERIVHSDLKPANFLLVRGFLK
                     LIDFGIAKAINSDTTNIQRDSQVGTLSYMSPEAFMCNESDENGNTIKCGRPSDIWSLG
                     CILYQMVYGRTPFADYKTFWAKFKVITDPNHEITYNQLSNPWLIDLMKKCLAWDRNQR
                     WRIPELLQHPFLAPPIPHEPQVKTIKLFSLIAESCGSDDDKANSMISQLEQLLSNPAP
                     LPRNDVLDSRDQNQQLLSRVSELCIQLQDRLQGLEGEN"
     gene            21802. .22921
                     /gene="T32E8.6"
                     /note="similar to putative SWH1 protein GB:AAC20736
                     [Arabidopsis thaliana]; Pfam HMM hit: PH (pleckstrin
                     homology) domain"
     mRNA            join(<21802. .22203,22526. .>22921)
                     /gene="T32E8.6"
     CDS             join(21802. .22203,22526. .22921)
                     /gene="T32E8.6"
                     /codon_start=1
                     /product="hypothetical protein; 21802-22921"
                     /protein_id="AAG51621.1"
                     /db_xref="GI:12323289"
                     /translation="MSSPLAAKSNLTTTRSASLNIPHSTRTNHNHDLRYSLSAGPRTN
                     EDRPASGNGVAGILYKWVNYGQGWKRRWFVLQDGVLSYYRIHGPDKISLSVEMDRRSK
                     LIGGESLRFICRHSKRGDVHSPGKPLGQIHLKVSSIGQSISDGKRFTVFTGTKSLHLR
                     AATSEDRASWIEALKAVKETFPRMSNEELMASTTNVSVSTDKLRQRLMEEEVDETIIK
                     DCEDIMKNNFLALHDEVMSLKRYQYHLIDSLKNVSNSPIRPSNQSFS"
     gene            27895. .31218

  Query Match          100.0%;  Score 2149;  DB 8;  Length 82454;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42450 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 42391

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42390 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 42331

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42330 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 42271

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42270 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 42211

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42210 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 42151

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42150 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 42091

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42090 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 42031

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42030 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 41971

Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41970 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 41911

Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41910 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 41851

Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41850 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 41791

Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41790 AGAACAATTTAATAAAGTGGATAAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 41731

Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41730 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 41671

Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41670 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 41611

Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41610 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 41551

Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41550 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 41491

Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41490 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 41431

Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41430 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 41371

Qy       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41370 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 41311

Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41310 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 41251

Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41250 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 41191

Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41190 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 41131

Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
```

```
Db      41130 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 41071

Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Db      41070 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 41011

Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Db      41010 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 40951

Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Db      40950 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 40891

Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Db      40890 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 40831

Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Db      40830 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 40771

Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Db      40770 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 40711

Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Db      40710 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 40651

Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Db      40650 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 40591

Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Db      40590 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 40531

Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Db      40530 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 40471

Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Db      40470 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 40411

Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Db      40410 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 40351

Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149

Db      40350 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 40302
```

```
RESULT 7
AF242313
LOCUS       AF242313               1741 bp    DNA     linear   PLN 06-AUG-2000
DEFINITION  Arabidopsis thaliana unknown sequence.
ACCESSION   AF242313
VERSION     AF242313.1  GI:9719364
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1741)
  AUTHORS   Mollier,P., Hoffmann,B., Pelletier,G. and Chwetzoff,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Genetique, INRA, Route de St Cyr,
            Versailles 78026, France
```

```
FEATURES             Location/Qualifiers
     source          1. .1741
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="NIA1-agp64"
                     /ecotype="Wassilewskija"
     misc_feature    <1. .>1741
                     /note="contains root-specific promoter"
ORIGIN
```

```
  Query Match             80.6%;  Score 1733;  DB 8;  Length 1741;
  Best Local Similarity   99.7%;  Pred. No. 4.8e-292;
  Matches 1736;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        409 AATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGG 468

Db          1 AATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGG 60

Qy        469 TTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAAT 528

Db         61 TTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAAT 120

Qy        529 TTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAG 588

Db        121 TTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAG 180

Qy        589 GGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGG 648

Db        181 GGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGG 240

Qy        649 AAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAA 708

Db        241 AAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAA 300

Qy        709 GAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGA 768

Db        301 GAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGA 360

Qy        769 CGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTC 828

Db        361 CGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTC 420

Qy        829 AAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTT 888

Db        421 AAAACCATATCCAATTTTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTT 480

Qy        889 GTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCT 948

Db        481 GTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCT 540

Qy        949 ATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCA 1008

Db        541 ATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCA 600

Qy       1009 AAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAA 1068

Db        601 AAATTTGATTAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAA 660

Qy       1069 ATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATG 1128

Db        661 ATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATG 720

Qy       1129 ATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTT 1188

Db        721 ATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTT 780

Qy       1189 GTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTA 1248

Db        781 GTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTA 840

Qy       1249 TATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTA 1308
```

```
Db        841 TATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTA 900

Qy       1309 GTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGT 1368

Db        901 GTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGT 960

Qy       1369 AATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTT 1428

Db        961 AATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTT 1020

Qy       1429 GCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTT 1488

Db       1021 GCCACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTT 1080

Qy       1489 GTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAA 1548

Db       1081 GTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAA 1140

Qy       1549 GATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCA 1608

Db       1141 GATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCA 1200

Qy       1609 TATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCC 1668

Db       1201 TATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCC 1260

Qy       1669 CTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGC 1728

Db       1261 CTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGC 1320

Qy       1729 ACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATT 1788

Db       1321 ACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATT 1380

Qy       1789 CAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAA 1848

Db       1381 CAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAA 1440

Qy       1849 TATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATA 1908

Db       1441 TATTCTGTTTAACTCACTGACGAATTTTTTATTTTTTCAAAGAAGAACATTTTTGATATA 1500

Qy       1909 AATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTT 1968

Db       1501 AATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTT 1560

Qy       1969 TTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAA 2028

Db       1561 TTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAA 1620

Qy       2029 CATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGAT 2088

Db       1621 CATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGAT 1680

Qy       2089 GACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCA 2148

Db       1681 GACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCA 1740

Qy       2149 G 2149

Db       1741 G 1741
```

```
RESULT 8
ATH553603/c
LOCUS       ATH553603               317 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            362F10.
ACCESSION   AJ553603
VERSION     AJ553603.1  GI:29369764
KEYWORDS    left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
  MEDLINE   22363535
   PUBMED   12446565
REFERENCE   2  (bases 1 to 317)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Wassillewskija"
                     /db_xref="taxon:3702"
                     /clone="362F10"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
     misc_feature    1. .317
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN
```

```
  Query Match             11.7%;  Score 251.8;  DB 8;  Length 317;
  Best Local Similarity   99.2%;  Pred. No. 4.6e-34;
  Matches  253;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1753 AAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGCATCGTTTCAAAT 1812

Db        317 AAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGCATCGTTTCAAAT 258

Qy       1813 TTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAACTCACTGACGAA 1872

Db        257 TTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAACTCACTGACGAA 198

Qy       1873 TTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACCACCGG 1932

Db        197 TTTTTTATTTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACCACCGG 138

Qy       1933 TTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCATTAACGACCTACA 1992

Db        137 TTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCATTAACGACCTACA 78

Qy       1993 TTTGATCCCTCATTA 2007

Db         77 TTTGATCCCTCATAA 63
```

```
RESULT 9
AL773135
LOCUS       AL773135                295 bp    DNA     linear   STS 11-JUN-2003
DEFINITION  Arabidopsis thaliana transposon insertion STS SM_3.1249, sequence
            tagged site.
ACCESSION   AL773135
```

```
VERSION     AL773135.1  GI:21535322
KEYWORDS    STS; STS, sequence tagged site.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B.,
            Murphy,G., Langham,S., LeGrys,C., Jones,J.D.G. and Bevan,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 295)
  AUTHORS   Clarke,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-2002) Clarke J.H., John Innes Centre, Colney
            Lane, Norwich, NR4 7UJ, UK
COMMENT     AT denotes an activation tag dissociation transposon within a
            single line, ET an enhancer trap dissociation transposon, GT a gene
            trap dissociation transposon, MT a mis-expression enhancer trap
            dissociation transposon, SM a defective suppressor mutator
            transposon. _3 denotes a sequence derived from the 3'end of the
            transposon, _5 denotes a sequence derived from the 5'end of the
            transposon BBSRC GARNet, ATIS project
            On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
            code: N55909.
FEATURES             Location/Qualifiers
     source          1. .295
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia-0 NASC stock code N1092"
                     /db_xref="taxon:3702"
                     /clone="AC012193"
                     /note="Derived from superpool 12.77 NASC code Unknown"
     STS             1. .295
                     /standard_name="SM_3.1249"
ORIGIN

  Query Match             6.1%;  Score 132;  DB 11;  Length 295;
  Best Local Similarity   76.4%;  Pred. No. 3.3e-13;
  Matches  162;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        321 ACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAA 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAA 60

Qy        381 AAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTCTCGAATTGCCTCAACAATTT 440
              |||||||| | || || | ||| ||  |||  || ||| | | | || || ||||| |||
Db         61 AAAAAAAAGGGGGAATCCGGGTTTAACCAAACCCCCGGCCCCAATTTCCCCCAACATTTT 120

Qy        441 CTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATTCTTAATAACATACTCTATAT 500
              || ||| |||  || || | | ||| |||||| |||| | || || | ||  | |||| |
Db        121 CTGGGAAAAAGGGGCCGGGCCCCTTGGGTTTTTTTTTTTCCTAAAAACCACCCCCTATTT 180

Qy        501 TTTAAACACTTCGATGTCTCGCTTAAATTTCG 532
              || || | || | | | | |  | || ||| |
Db        181 TTAAACCCCTCCAAGGCCCCCTTAAATTTTTG 212


RESULT 10
AC007289/c
LOCUS       AC007289               43982 bp    DNA     linear   PLN 27-FEB-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone F16J10 map RNS1, complete
            sequence.
ACCESSION   AC007289
VERSION     AC007289.9  GI:20198243
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 43982)
  AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
            Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
            Nierman,W.C., Fraser,C.M. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43982)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 43982)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598749.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .43982
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="RNS1"
                     /clone="F16J10"
     misc_feature    1. .5724
                     /note="overlap with BAC clone T20G20
                     (AC006220:104629. .110352)."
     misc_feature    1
                     /note="17532 nt  before this point were not included in
                     the submitted sequence, due to overlap with another BAC"
     repeat_region   complement(1434. .1472)
                     /rpt_family="AT_rich"
     repeat_region   complement(1907. .1945)
                     /rpt_family="AT_rich"
     gene            complement(2734. .3407)
                     /gene="At2g05435"
                     /note="synonym: F16J10.1; pseudogene"
                     /pseudo
     repeat_region   complement(3455. .3489)
                     /rpt_family="(TA)n"
     repeat_region   complement(3808. .3836)
                     /rpt_family="AT_rich"
     repeat_region   complement(3859. .3914)
                     /rpt_family="Rf:ATREP9|ATREP9 ATREP9 repeat - a
                     consensus."
     repeat_region   complement(3978. .4108)
                     /rpt_family="AT_rich"
     repeat_region   complement(4754. .4787)
                     /rpt_family="AT_rich"
     gene            complement(5500. .6147)
                     /gene="At2g05430"
                     /note="synonym: F16J10.2; predicted by genefinder"
     mRNA            complement(join(<5500. .5949,6031. .>6147))
                     /gene="At2g05430"
     CDS             complement(join(5500. .5949,6031. .6147))
                     /gene="At2g05430"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAM15480.1"
                     /db_xref="GI:20198245"
                     /translation="MSKISESKSTLSHHGTAYDEDLDNDEIVNIKGFWIFRSKLGLAK
                     RVFEKHPETTTKFCLKSEFAKETYLTALLDLIDMINMLPQQSLFEAELKEAENTILDL
                     EAAGFKLDWLKRKLEEIRVTKKKAKNRTTRMRELDRKIQKHLEELSVLQEEMKKEQFE
                     AMCDKPEYYLDSLCLNEVTCDSSEPITK"
     gene            complement(7260. .8700)
                     /gene="At2g05420"
                     /note="synonym: F16J10.3; predicted by genefinder"
     mRNA            complement(join(<7260. .7640,7857. .8077,8329. .8510,
                     8591. .>8700))
                     /gene="At2g05420"
```

```
CDS             complement(join(7260. .7640,7857. .8077,8329. .8510,
                8591. .8700))
                /gene="At2g05420"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAD32895.1"
                /db_xref="GI:4914358"
                /translation="MGKQVDSKTITWVIENFSSLQSASIHSDQFVVGDCKWRLKAYPK
                GNEKATYLAYRANNLALYLNVANSKSFPIGWTRHTKFSLTLVNQKSEKLSKLTESQHW
                FDHKSTSRGFPAMIPLTNLHTNEGFLVNGELTLVAKVEVLEVVGKLDVSKKSSPVMET
                IDVNGFQVLPQIESVNRLFAKHLDIASKFRPKKPYMKTAYMNVLLSLTKTLCQSPQDL
                SNDDISGAGAALTYLREAGFKLDWLEKKHGEIKEKKKKEEASLKRLQEMEKQIFNEAQ
                IYKEKVLAARAPLSLNEDNVF"
repeat_region   9262. .9302
                /rpt_family="AT_rich"
repeat_region   complement(9910. .9970)
                /rpt_family="AT_rich"
repeat_region   complement(10000. .10019)
                /rpt_family="(TAAA)n"
repeat_region   10066. .10101
                /rpt_family="AT_rich"
gene            complement(14048. .15111)
                /gene="At2g05410"
                /note="synonym: F16J10.4; predicted by genscan"
mRNA            complement(join(<14048. .14464,14624. .14847,14955. .>15111))
                /gene="At2g05410"
CDS             complement(join(14048. .14464,14624. .14847,14955. .15111))
                /gene="At2g05410"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAD32896.1"
                /db_xref="GI:4914359"
                /translation="MAYPKGINKAHDSFSLFLNVPDNESLPTGWRRHAKVSFSLVNQG
                SEKLSQRKVTQHWFVQKAFTWGFPVMITHTELNAKMGFLVNGELKVVAKIEVLEVVGK
                LDVSKESSPIMKTIDVNGFQVLPSQVDSVKRLFEKNLDIVSKFRLKNPYLKTACMNLL
                LSLTETLCQSPQELSNDGLSDAGVALAYLIETGLKLDWLEKKLDELKEKKKKEESCLV
                RLREMDEQLQPFKKRCLDIEDQISKEKEELLAAREPLSLYDDIDNNV"
repeat_region   complement(15873. .15972)
                /rpt_family="AT_rich"
gene            complement(16262. .17096)
                /gene="At2g05400"
                /note="synonym: F16J10.5; predicted by genscan"
mRNA            complement(join(<16262. .16672,16854. .>17096))
                /gene="At2g05400"
CDS             complement(join(16262. .16672,16854. .17096))
                /gene="At2g05400"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAD32897.1"
                /db_xref="GI:4914360"
                /translation="MFCDSAETQQWFDKNAPDWGFTEMISLTELHAEEGLLRNGELTV
                VAKVEVLEVVGKLDVSEESSPVIETIDVNGFQVLPSQVEYAKSLFERHLDIASKFRPK
                NPYLKTAYMNVLLSLTQTICQSPQELSNDDLSDAGAALAYLREAGFELDWLEKKLNEV
                KEKKKKEEACLAEIQDMDEHVKPLKKKYLDLEAQIDKKKAELLAARAPLSLNDDNVV"
gene            complement(17875. .21892)
                /gene="At2g05390"
                /note="synonym: F16J10.6"
mRNA            complement(join(<17875. .19791,19831. .20615,20671. .>21892))
                /gene="At2g05390"
CDS             complement(join(17875. .19791,19831. .20615,20671. .21892))
                /gene="At2g05390"
                /codon_start=1
                /product="putative retroelement pol polyprotein"
                /protein_id="AAD32898.1"
                /db_xref="GI:4914361"
                /translation="MEATLRVHKVWETIDPGSDDMEKNDMARALLFQSVPESTILQVG
                KHKTSKAMWEAIKTRNLGAERVKEAKLQTLMAEFDRLNMKDNETIDEFVGRISEISTK
                SESLGEEIEESKIVKKFLKSLPRKKYIHIIAALEQILDLNTTGFEDIVGRMKTYEDRV
                CDEDDSPEEQGKLMYANSESSYDTRGGRGRGRGRSSGRGRGGYGYQQRDKSKVICYRC
                DKTGHYASECLDRLLKLIKAQEQQQNNEDDDEIESLMMHEVVYLNERSVKPKEFEACS
                DNSWYLDNGASNHMTGNLQWFSKLNEMITGKVRFGDDSRIDIKGKGSIVLITKGGIRK
                TLTDVYFIPDLKSNIISLGQATEAGCDVRMKDDQLTLHDREGCLLLRATRSRNRLYKV
                DLNVENVKCLQLEAATMVRKELVIGISNIPKEKETCGSCLLGKQARQPFPKATTYRAS
                QVLELVHGDLCGPITQSTTAKKRYILVLIDDHTRYMWSMLLKEKSEAFEKFRDFKTKV
                EQESGVKIKTFRTDKGGEFVSQEFQDFCAKEGINRHLTAPYTPQQNGVVERRNRTLLG
                MTRSILKHMKMPNYLWGEAVRHSTYIINRVGTRSLQNQTPYEVFKQRKPNVEHLRVFG
                CIGYAKIEGPHLRKLDDRSKMLVYLGTEPGSKAYRLLDPTNRKIIKWNNSDSETRDIS
                GTFSLTLGEFGNNGIQESDDIETEKNGEESENSHEEEGENEHNEQEQIDAEETQPSHA
                TPLPTLRRSTRQVGKPNYLDDYVLMAEIEGEQVLLAINDEPWDFKEANKLKEWRDACK
                EEILSIEKNKTWSLIDLPVRRKVIGLKWVFKIKRNSDGSINKYKARLVAKGYVQRHGI
                DYDEVFAHVARIETIRVIIALAASNGWEVHHLDVKTAFLHGELREDVYVTQPEGFTNK
                DNEGKVYKLHKALYGLKQAPRAWNTKLNKILQELNFVKCSKEPSVYRRQEEKKLLIVA
                IYVDDLLVTGSSLDLILCFKKDMAGKFEMSDLGQLTYYLGIEVLHRKNGIILRQERYA
                MKIIEEAGMSNCNPVLIPMAAGLELCKAQEEKCITERDYRRMIGCLRYIVHTRPDLSY
                CVGVLSRYLQQPRESHGNALKQVLRYLKGTMSHGLYLKRGFKSGLVGYSDSSHSADLD
                DGKSTAGHIFYLHQCPITWCSQKQQVVALSSCEAEFMAATEAAKQAIWLQDLFAEVCG
                TTSEKVMIRVDNKSAIALTKNLVFHGRSKHIHRRYHFIRECVENNLVEVDHVPGVEQR
                ADILTKPLGRIKFREMRELVGVQCVLEDDFKLKGEFVGISMKSIN"
repeat_region   complement(22479. .22502)
                /rpt_family="AT_rich"
gene            complement(24570. .25781)
                /gene="At2g05380"
                /note="synonym: F16J10.7; supported by cDNA:
                gi_15146251_gb_AY049267 1_"
mRNA            complement(join(<24570. .25058,25678. .>25781))
                /gene="At2g05380"
CDS             complement(join(24781. .25058,25678. .25750))
                /gene="At2g05380"
                /codon_start=1
                /product="expressed protein"
                /protein_id="AAD32899.2"
                /db_xref="GI:20198244"
                /translation="MASKTLLLLGLFAFLFIVSEMAAAGTVKSESEETVKPEQHGGGF
                GDNGGGRYQGGGGHGGHGGGGYQGGGGRYQGGGGRQGGGGSYCRHGCCYKGYHGCSRC
                CSYAGEAVQTQSGH"
repeat_region   complement(24878. .24990)
                /rpt_family="(GGA)n"
repeat_region   complement(25118. .25153)
                /rpt_family="AT_rich"
repeat_region   complement(26299. .26329)
                /rpt_family="(A)n"
repeat_region   complement(28917. .28939)
                /rpt_family="AT_rich"
repeat_region   28929. .29834
```

Query Match 4.4%; Score 93.8; DB 8; Length 43982;
Best Local Similarity 89.4%; Pred. No. 7.8e-07;
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy         74 TTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAATCT 133
              || || ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      34629 TTTTAACTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAACCT 34570

Qy        134 TTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACAAGTTCT 186
              |||||||| ||| |||||||||||||| |||||||||||||||  | ||  | |
Db      34569 TTCATTTCTCGAAATGATATTTACATTCTTATCAAAAAAAAAAAAAAAAGTTT 34517
```

```
RESULT 11
PFA929353
LOCUS       PFA929353             343050 bp    DNA     linear   INV 29-JAN-2003
DEFINITION  Plasmodium falciparum strain 3D7, chromosome 5, segment 3/4.
ACCESSION   AL929353 AL844504
VERSION     AL929353.1  GI:23504644
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
```

```
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., Mclean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
  MEDLINE   22255708
   PUBMED   12368867
REFERENCE   2  (bases 1 to 343050)
  AUTHORS   Devlin,K., Baker,S., Davies,P., Mungal,K., Berriman,M., Pain,A.,
            Hall,N., Bowman,S., Churcher,C., Quail,M. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .343050
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="5"
     gene            complement(148. .1134)
                     /gene="PFE0795c"
     CDS             complement(148. .1134)
                     /gene="PFE0795c"
                     /note="Similar to Mus musculus nif-like protein nifL
                     TR:CAC69078 (EMBL:AJ344340) (348 aa) fasta scores: E():
                     7.1e-16, 36.310% id in 168 aa
                     Similar to Dictyostelium discoideum development protein
                     Dg1148 dg1148 TR:Q9XYL0 (EMBL:AF111941) (306 aa) fasta
                     scores: E(): 2.6e-16, 37.267% id in 161 aa"
                     /codon_start=1
                     /product="nif-like protein, putative"
                     /protein_id="CAD51523.1"
                     /db_xref="GI:23504645"
                     /db_xref="SPTREMBL:Q8I3U9"
                     /translation="MKKRKFERCLYQILKIVKFYMNIWIKIKKLIKYVLSPLFLRDSF
                     HFKTYRKRRICPTMTLVLDLDETLIYCTKKRKYHYQKEVDVLINGKYLPLYVCKRPYI
                     DLFFSSLYPFYEIIIFTTAIKSYADTVLNIIDVDHYIDKKFYREDCYEMNEKLYIKNL
                     TNIKKELSKIILIDDSNISGFQYPDNFFPIKKWQGDLNDNELLHLIPFFLNLRKLKDI
                     RSLCSFRLISQKDVSKLLCTSSSKHIKYLTDDNSKYLYRHSLAFQAINYLKIFMNKFR
                     YASSKKQWNELGKKINNKLKSYPFSLSKLKGSSDDEHKKKHRKIINNRKMLR"
     gene            join(3549. .6704,6817. .6915)
                     /gene="PFE0800w"
     CDS             join(3549. .6704,6817. .6915)
                     /gene="PFE0800w"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD51524.1"
                     /db_xref="GI:23504646"
                     /db_xref="SPTREMBL:Q8I3U8"
                     /translation="MRKVSIKFINSRLYYSRRYLSRNQFYSFFSCNKNEQRPIKSKND
                     DFLIDESYNNVEFDEYISPSFSFDINIDDDKWKKKEEIEYNKFIEKTKKDSINPSTFF
                     NTQKMTKHIKNTSGNSNYEKLQIHNKHIDNNISSNVVTSNNSTPHNVETFYEKENIYE
                     DDINIFLKNSSDDITLQDGGIIKKLIYLDSNYKSLHYENKKIGAYLNSILDLLERGST
                     DTHKDGVVKTRRGNEENEENGKNEENGKNEKNEENGKNEKNEKNEENGKNEKNEKNEENGK
                     NGKNGKNEKNEENGKNEQNEEDKVDKEYEADILYKDNKERTETPNERVGGGHIQYLQG
                     HIFNNIYEDPFSLNNLKYQKLVTNFTKFAGYYTNTPFKLIVENIQSLTYNNLVFTLNL
                     LDHLNMKNMMRDLSYEVIKMLNKKKKSDQEDILQNANINKKENKKINDIENNIKLCVH
                     YINILNFHSLYFCEFYDYLIEYINNMKNDVLCFFTLQCFTHSLRTKHYLDKIYFLCLK
                     KIHEFNYEELKCMIYCFHRFCKEYSKFYDLSIDIIKNKYIHLLNMNLNFVQLLLKITN
                     AFNKNDKYVELKGLIAEKIIMRVGNIQREEKKKKGQYLNNSTNVNTSVDMWTNVERHK
                     KKESSSRMCAPYDTTYLGANSCVSKVFMENSVHNTIYEDNIIKIIQCLKYLEYNKKNS
                     EHVKTCVNKIYEMINENMECLEKLDIEDIVYSIVCFSTYNKRLILYNNFLDIIYEKSN
                     ELMNAKNICLWIYPILSLCKISWYHRNYIELLFGYIKDNYILNRLSVFQVLKLLSSIV
                     KMNIYDEEVYKILIEKLYKEWDIIKKKLIDISTFLWCCAYINIIYKPLFDSAYNLIID
                     LINKDNVPIINNVMYKNCFVNITWSFIVANYHKTHTNFDQILNITFLKRNPNENEAFK
                     RLHQISDSCFKEIPKTLINITCLDTLYAYCTHEKCKLLRNNFIIYKKEKDAVKMKNKI
                     HDELINLLKNSFHISFDQYVEPYHNSPYIIDICLNQTSKIGVSIFGKEYLMRTLKKSS
                     WDFMNTGIVTLQMRILHAHGWKIIPINAGEWLQLNFDQKKNLLSEYFKQYSIQI"
     gene            join(7484. .14348,14522. .14859)
                     /gene="PFE0805w"
     CDS             join(7484. .14348,14522. .14859)
                     /gene="PFE0805w"
                     /EC_number="3.6.3.-"
                     /note="Similar to Plasmodium falciparum probable
                     cation-transporting ATPase 1 SW:ATX1_PLAFA (Q04956) (1956
                     aa) fasta scores: E(): 0, 97.919% id in 1970 aa
                     Large N- and C-terminal extensions
                     10 probable transmembrane helices predicted by TMHMM2.0 at
                     aa 50-72, 286-308, 334-356, 361-380, 708-730, 2173-2195,
                     2207-2229, 2263-2282, 2302-2324 and 2334-2356"
                     /codon_start=1
                     /product="cation-transporting ATPase 1"
                     /protein_id="CAD51525.1"
                     /db_xref="GI:23504647"
                     /db_xref="GOA:Q8I3U7"
                     /db_xref="SPTREMBL:Q8I3U7"
                     /translation="MYKNIFFPHDNIKKETYKVYYYHTNNLIHDEYRISYIECYKKKK
                     KKIVTILHYVFCILSLGLLPLILSWFPIFYFRLLHSACTLNECEKVLIITKDGNKYIK
                     KIKKIKFNHNVCYVMNDVFFSEDQSMSSMYNDKEQNKNNFNYFSRDYTFRDQNIMNER
                     KDNINLKDHIKLSGHINLKDNIKLSGHINLKDNIKLSGHINLKDNIKLSGHINFKDNI
                     NFKDSIYPCDEILHVDNIPLHNNLNINSHQPFDETTEYNKSRLQDNNRITYIFEFLDI
                     RYVYIKKKNIFVELYFNINYPFSYMHLMCTGLRNEKLINDRKILYGECNLNIKSDSFI
                     ILLFKEIMNPFFIFQIFAMIVWSLDNYIEYTISILFITSISIILELKNTIKNQKKIKN
                     MLNYTCPINVYRYNTSYIISSSELVPGDIYEIKNNMTIPCDTIILSGSVTMSEHMLTG
                     ESVPIHKEQLPFEGNAIINKNNKYESNDEKDDYLRIYNNHASINMIKRNHLIEETLGK
                     KDREYKSNTHDLCSMNKLCYINNTYDDVHMKNNKMDYNNNNNNKKKKKINNLNFVKGT
                     YINSNDLLYDDKIGVNIFEDDVNNMKHKFNQRNINYYNKDTNNLEYNNKHRYIYDCLL
                     KKVEAISQKNKIIYSNEDINKYMLYGGTYVLSLYNINKIKYNNKEENRILGLVIKTGF
                     ITTKGKIVNNILYHKKKELNLINDSYKFLIILIIYAFFSVFILLYITLSNNEYTNHII
                     IKCLDIITDAIPPALPTTLTVGISIAISRLKKKFSISCLCPHKINIAGQINTMVFDKT
                     GTLTENNLQFIGIITQNKKNKNMLSDFIHIKEMNTESYIHSKDDNMIHNKNSIISEYY
                     IKDNMKNLHTSSKKKSITKERSNFLVQTIKSCLLKDHYIKEKKKEYYTNNTYCNDLHI
                     NDSTCSSYLLNSETKDAYCEYYNIDHLCDINKKNMDINSKNELMGKYSKNELMGKYSK
                     NELMGKTIKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKYSKNELM
                     GKYSKNELMGKTIKNQVGVDTNIYHMNCDNDYNYDYPCDYNCNNCNDTYHRLEYHNIN
                     KDNSFNIPPEKNKSYNNISEHIKINYPLLFEALACCHTLSKVNNKIMGDVLEILMFNF
                     TNCDMLINNNSFIIKEKKKNCSYDFQKIDGDKNIGANDERCHLNNNLVSYNILKRFEF
                     QSRLQRMSVIVKSTYGNNNDDNNDDDNNNDDDNNDDNNNDDNNNDDNNDDNNNNNYYY
                     NIFCKGSPEKIKELCLKSKIPNNYDEILNKYTKQGMRILSISYKRVKSKNINLLNVKR
                     SFVESNLHFLGFLIFTNNMEKNAPDIIHNLQTSGCQCIMSTGDNVLTSIHVAKKCGII
                     NSNVESIIIGDVIPVVGKNNKQKKKLVWYNHKNDTYLKGHDKTCIDNEFTSIQSQMNS
                     DNICGDNICGDNICGDNIYGDNICGDNIYGDNINGDNINGDNINGDNIYGDNINGDNI
                     NGDNINTYDNIYGDNYNLDYCPTEYHKCTYNNSILYRNNFLYKKENKKDKNYKNISTL
                     YEHRTNDIQFDKLCDILINKDPRNVNIVLTGKAFIFLKKKFYSFHLPYYEECKNIVHY
                     IMKKKHKKIKNIINNHNSNLYYHYNIIDTFVKRMNKEYMCFNKLLYKIQQKLLYNLIH
                     NLYKKKKYMNNYYDIDEVHLIGNNNNNNKNNSKEKKLPLKNKMKHIRKNESNDNITFN
                     TYTSNNIYLSKYKYVHHKNYYYPDSCTNLRKKKNSLFYNLKKYIYYEKKKYLQHCLLK
                     HDNYKKVELPRIKDINYSYQMESIKTRNFIHSLSEQFAIFSNLILSFYIIKNDDNNVY
                     NKNYIYNKNYIYNKNSICNKNYICNKNYIYNKNNIYNKNNIYNKKNILTHAKSVLLSG
                     SSKKFLKFFSNIIRRHKLKEKKNKKNIKRYKMNHVNNTSKGHIILNMCTHGFKKDYSS
                     LKNKYRIVNNKGYMLKNDNVYDRHMYNLTDMYRGTQYGCSKKKNKNIYMNNNNNILKN
                     KINRFLEHLLVDKCKRNICHKYTDIKNIKLSIYEYILRTCTVYARMKPKDKSDLILSL
                     KKLPNNSYVGMCGDGANDCLALSCADIGISLCNNNESSICSSFTSNKLCLHSIVHILI
                     EGRASLVNSFQLFKFISLYSIMQCSQVLILYSISNKLTDNQYIFIDIVTILPLSIFMC
                     WTSASEKLSKNIPIGKLFSFPILISIYGQIIIQLFFVMISLVLLMNLSFYKYDKNKVM
                     KEKSDDTYLYKAQNTLLYILCSFQNLFMCISLNIKNEWRKSVFTNIAFIIWMSFLLLM
                     NTCITFFSSEIFLVGWIIDLLKQYLSLITFPFYFRIFLFFILFFNFLCSYSYEKYIIK
                     YFEKREIQKKYNHNHMSSIFCSPQELDTTNVFVTCN"
     gene            complement(join(15807. .15912,16021. .16198,16441. .16612))
                     /gene="PFE0810c"
     CDS             complement(join(15807. .15912,16021. .16198,16441. .16612))
                     /gene="PFE0810c"
                     /note="Similar to Arabidopsis thaliana putative 40s
                     ribosomal protein s14 f24k9.19 TR:Q9CAX6 (EMBL:AC008153)
                     (150 aa) fasta scores: E(): 2.9e-44, 81.457% id in 151 aa"
                     /codon_start=1
```

```
                     /product="40S ribosomal subunit protein S14, putative"
                     /protein_id="CAD51526.1"
                     /db_xref="GI:23504648"
                     /db_xref="GOA:Q8I3U6"
                     /db_xref="SPTREMBL:Q8I3U6"
                     /translation="MASKKVKTPQPETAIVSGPQPKEGELVFGVAHIFASFNDTFIHV
                     TDLSGRETLVRITGGMKVKADRDESSPYAAMMAAQDVAARLKELGVTAIHIKLRASGG
                     TKSKTPGPGAQSALRALARSGLKIGRIEDVTPIPTDSTRKKSGRRGRRL"
     gene            18908. .20953
                     /gene="PFE0815w"
     CDS             18908. .20953
                     /gene="PFE0815w"
                     /note="Weak similarity to several tRNA pseudouridine
                     synthases, incl. Staphylococcus aureus tRNA pseudouridine
                     synthase a truA TR:BAB45147 (EMBL:AP003364) (267 aa) fasta
                     scores: E(): 0.00011, 30.952% id in 126 aa
                     Signal peptide predicted by SignalP 2.0 HMM (Signal
                     peptide probabilty 0.925, signal anchor probability 0.071)
                     with cleavage site probability 0.497 between residues 24
                     and 25
                     1 probable transmembrane helix predicted by TMHMM2.0 at aa
                     5-24"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD51527.1"
                     /db_xref="GI:23504649"
                     /db_xref="SPTREMBL:Q8I3U5"
                     /translation="MNKFFPLLVCVSLIFFFLKLSINAKKIFKKNIKNKPFILKKSNN
                     INGIYFIKSKLNKIENKKNKNKKMKLFCDHNDNVNRMEQQNNSVNVSKDNSSFIFSNR
                     KKLTNLLLIVNYDGTNYNGWTGLENSSEVYLNALQNYKNKKKTERQKNIERKKYSTVQ
                     NNILDCILKLHGYKYIHNNNNNEHICNSYKVGDTDNYNNAHHNDYNQNNVNGIINNKP
                     FEFIGVSRTDKGVHAKEYICQYISYEKEPPCDGDMEHIKRSLNSLLNKDIKILAVLKS
                     PHDHFNIRFHNSGKIYTYNLDIRNPSQPLERNYAWQLYDDPRFFFLSKKTNKHKKEGT
                     PNSIHNAQQTCTTPGDKNGRKALYIANKKEEDLKYIYDDELSLLFGENYIPNGTYDDE
                     NLSDEKQNCLDCENISTNHISSSNDPPNDNICLSNDLPNDNICLSNDLPNDNICLSND
                     LPNDNICLSNDLPNDNICLSNDLPNDNISPSNDLPNDNISPSNDLPN
                     DNISPSNDPPSFYHKNMKSSYNDISHSMKIINKEKEIRSVIPCDINKIKECAKLFIGH
                     HNFECFRGTLKGTEKLRKINTFCTIHFLDVYELKNNLYQFVIQGDRFLYHMIRIIVGT
```

Query Match 3.5%; Score 76.2; DB 3; Length 343050;
Best Local Similarity 42.2%; Pred. No. 0.0007;
Matches 495; Conservative 0; Mismatches 673; Indels 5; Gaps 1;

```
Qy       138 TTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACAAGTTCTACCAAAGCACA 197
Db    327206 TTTATTTAATTTTTTTTTTTATTATTTTTTATATTATTGTTTCATATTATACGCAAAAAAA 327265

Qy       198 GGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAATCTTATGATTTCCCTTT 257
Db    327266 AAGAAAAGAAATATATTATTTTATATTCATAATTTTAAGGCAATTATATATTATCTTTAT 327325

Qy       258 TCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTGTCATTAACTTCCCACTC 317
Db    327326 ATATATATATACATATATATATAGATATTTATTTATTTATTTATTTTAAAATAAAAGA 327385

Qy       318 TATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAA 377
Db    327386 ATAACTTAATTTTATTGTATTCTTATTAATATAAATAAACTATATTCCTTATATATTAAA 327445

Qy       378 AAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTCTCGAATTGCCTCAACAA 437
Db    327446 ATTAATAATAGGAATGTTATAATAATATTATTATTATACATTATAAGAATTCAGTATGTA 327505

Qy       438 TTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATTCTTAATAACATACTCTA 497
Db    327506 TTATAATTAAAAATATGAAAATTTTATTTTGCATTTCATTATTTTCTTTTATATAGAGGT 327565

Qy       498 TATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCCTAAATTTCTCTAATCAT 557
Db    327566 TTTGTTTACATGAAATTATTGAAAACTAATAATTAAATATTTATTAGTATAATTAAATTT 327625

Qy       558 AAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAGGGCACGTAGTTACCTTT 617
Db    327626 AAAAAATAAAAATAAAGAAAATAATGAAAAAAAAGGAAAAGTGAAAAGGAAAAAAAAAAA 327685

Qy       618 AAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAAAGAACAATTTAATAAAG 677
Db    327686 AATAAAAGAAAAAAGAAAAGAAAAGAAAAGAAAAAGAAAGAAAAGAAAGAAAAAAAA 327745

Qy       678 TGGATAAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGGCCGCGACTCGTAACAAG 737
Db    327746 AAGAAAATAAATGAAAATAAAAAAAAAATAATAAAAGAAAAAAGAAAAGAAAAGAAAAAA 327805

Qy       738 GGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATTTTTTCTTTTTCTTTTCC 797
Db    327806 TAAAGGAAAAAAAGAAAAGAAAAAAGAACAAAACAAAAAGAACAAAAACATTTTTGTTGG 327865

Qy       798 TAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCCAATTGTTCTGCCTATTA 857
Db    327866 GTATTAAAAATATTTCATATATAATTATATATTATGTATATTCCAATTATATATATATAA 327925

Qy       858 TTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGTGATATATAATCTAACTT 917
Db    327926 TAATTATATATATAGTTTTATTTAGTATATATAATTTATATTATATTAAATTTATTTTTT 327985

Qy       918 AAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTGACCTCAGTCCCTAGTTA 977
Db    327986 TAATAAAATAAAAAAAAATATATACATATATACATAAATGATAACATACATATAATTTAT 328045

Qy       978 GCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGCAACT 1037
Db    328046 TATATATTCCTTTCATTGAATATGCATTAAATATATTATTTATATTAAATTTTTTATTGA 328105

Qy      1038 AACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCGTATTAAA 1097
Db    328106 ATGTATATATATATATATATATATATATTATATTAAATAATATATATATATATATATA 328165

Qy      1098 TATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGT 1157
Db    328166 TATTATATTAAATAATATATATATATATATATATTAAACATTTTTATTCTTTAAGTATAT 328225

Qy      1158 TATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTATA 1217
Db    328226 TTTATATTCAGACTTATATATTTTTATATTAATAAT-----ATTATTTATTATATATATT 328280

Qy      1218 TATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGTTG 1277
Db    328281 ATTTGGTACTTTATTTATATATTATATGCATAATATTAATATATAATAATATAAATAATA 328340

Qy      1278 TCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGT 1310
Db    328341 TATTTCAAGAAATTCTACTATATTATTAATAAT 328373
```

```
RESULT 12
AL731547/c
LOCUS       AL731547              175842 bp    DNA     linear   PRI 27-SEP-2002
DEFINITION  Human DNA sequence from clone RP11-565H13 on chromosome 10,
            complete sequence.
ACCESSION   AL731547 AC073586
VERSION     AL731547.9  GI:23380932
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175842)
  AUTHORS   Wray,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 30, 2002 this sequence version replaced gi:21694438.
            Draft Sequence Produced by Genome Therapeutics Corp, 100 Beaver
            Street, Waltham, MA 02453, USA
            http://www.genomecorp.com
```

```
              -------------- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquery@sanger.ac.uk
              --------------
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence was finished as follows unless otherwise noted: all
              regions were either double-stranded or sequenced with an alternate
              chemistry or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by at least
              one plasmid subclone or more than one M13 subclone; and the
              assembly was confirmed by restriction digest. The following
              abbreviations are used to associate primary accession numbers given
              in the feature table with their source databases: Em:, EMBL; Sw:,
              SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
              database can be found at
              http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
              was generated from part of bacterial clone contigs of human
              chromosome 10, constructed by the Sanger Centre Chromosome 10
              Mapping Group.  Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr10
              RP11-565H13 is from the library RPCI-11.2 constructed by the group
              of Pieter de Jong. For further details see
              http://www.chori.org/bacpac/home.htm
              VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..175842
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-565H13"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match              3.3%;  Score 71;  DB 9;  Length 175842;
  Best Local Similarity   46.1%;  Pred. No. 0.0061;
  Matches  310;  Conservative    0;  Mismatches  360;  Indels    3;  Gaps    2;

Qy        818 TTACAAGGGTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCC 877
Db     162378 TAAAAATATTAAATATTATATATTATATTTAAATATTTTATATTTATATATTATATTTAA 162319

Qy        878 TCTTGTGCTTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATAT 937
Db     162318 ATATTTTCTATTTATATATTATATATAAATATTTTCTATTTATATATTATATTTAAATAT 162259

Qy        938 ATGTCCTACCTATGTT-TCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGA 996
Db     162258 TTTCTATTTATATATTATATATAAATATTTTCTATTTATATATTATATATAAATATTTTC 162199

Qy        997 AAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATT 1056
Db     162198 TATTTATATATTATATATAAATATTTTCTATTTATATATTATATATAAATATTTTCTATT 162139

Qy       1057 GAGCTATGTTAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGA 1116
Db     162138 TATATATATTATATATAAATATTTTCTATTTATATATTATATATAAATATTTTCTATT 162079

Qy       1117 TAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATT 1176
Db     162078 TATATATAATATATAAATATTTGTTTATATATATTATATATATTTTATATTTATATATT 162019

Qy       1177 ACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATA 1236
Db     162018 ATATATATTTATATATAAATATATTTTATATATATTATATTTTATATAAATATATTTTA 161959

Qy       1237 TATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTT 1296
Db     161958 TATAAATATTTATATATAATATATTTTATATATAATATATTATATAATATATTATATATT 161899

Qy       1297 TTTTGTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAGA--AAAATCCAAATAGTTTAA 1354
Db     161898 ATATATAACATATTTATATATTATATATAAAATATATATAATATATTTTATATATTATAT 161839

Qy       1355 GGAAATTTTTGTGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCA 1414
Db     161838 TTAATATAAATATATTTTATATATATTATACTTATCTATAAATATATTTTTATATAATAA 161779

Qy       1415 TCATATTCCTCTTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAA 1474
Db     161778 TAAAATATATTTATATATAAATATAGTTATATATAATAATAAAATATATTTATATATAAA 161719

Qy       1475 AGTCGTCATTTGT 1487
Db     161718 TATAGTTATATAT 161706


RESULT 13
AE001398
LOCUS       AE001398               14867 bp    DNA     linear   INV 04-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 2 section 35 of 73 of the
            complete sequence.
ACCESSION   AE001398 AE001362
VERSION     AE001398.1  GI:3845197
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 14867)
  AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
            Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,
            Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M.,
            Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
            Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
  TITLE     Chromosome 2 sequence of the human malaria parasite Plasmodium
            falciparum
  JOURNAL   Science 282 (5391), 1126-1132 (1998)
  MEDLINE   99021743
   PUBMED   9804551
REFERENCE   2  (bases 1 to 14867)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419, 498-511 (2002)
REFERENCE   3  (bases 1 to 14867)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..14867
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="2"
     repeat_region   complement(<1..20)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(65..137)
```

```
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(178. .211)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        307. .327
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
repeat_region        complement(328. .348)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        complement(378. .423)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(479. .503)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(513. .539)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
repeat_region        577. .633
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
repeat_region        complement(714. .747)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        841. .867
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
repeat_region        complement(888. .918)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(1006. .1049)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(1075. .1120)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(1140. .1283)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        1294. .1346
                     /rpt_type=tandem
repeat_region        1298. .1346
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
repeat_region        complement(1423. .1531)
                     /note="AT_rich"
                     /rpt_type=tandem
gene                 complement(<1570. .>2424)
                     /locus_tag="PFB0490c"
mRNA                 complement(<1570. .>2424)
                     /locus_tag="PFB0490c"
CDS                  complement(1570. .2424)
                     /locus_tag="PFB0490c"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAC71887.1"
                     /db_xref="GI:3845198"
                     /translation="MKEKNEKIMDYLSCPLDDVVDREKKSGKNSLLKSSSTKKSDYKK
                     SSIFSKKRDSHKKGSSFRGRRSGFINRKSGSFKKPYYNNRLINKNYNNYKGRNFHNGR
                     DNFKGRTGSFGSRVFDNRKGSFKKRFISNRNKSSVKSYRGNGSNKMGRKSFNKAPTSR
                     TVVTKRLNNYKTVSAPVKKFNNLNISLYRKNRTFALNTKRSKPVGTIKSSVPRKRIKK
                     GLKKGSLKSKTRKSTSGSKFKPLNKYFLSKIKIVTSLNKIPSPLKEQKNTEVNLPESL
                     NNATTKKN"
repeat_region        complement(2150. .2188)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(2446. .2468)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        2504. .2656
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
repeat_region        2684. .2739
                     /rpt_type=tandem
repeat_region        2693. .2736
                     /rpt_type=tandem
repeat_region        complement(2697. .2732)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        complement(2757. .2845)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(2856. .2880)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        2884. .2953
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
repeat_region        2888. .2924
                     /rpt_type=tandem
repeat_region        complement(2977. .3167)
                     /rpt_type=tandem
                     /rpt_unit="(TAAA)n"
repeat_region        3047. .3132
                     /rpt_type=tandem
repeat_region        complement(3047. .3132)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        complement(3258. .3420)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(3450. .3514)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        complement(3532. .3559)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(3689. .3720)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(4024. .4114)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        4107. .4146
                     /rpt_type=tandem
repeat_region        complement(4115. .4146)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        4227. .4261
                     /rpt_type=tandem
repeat_region        complement(4227. .4257)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        complement(4348. .4386)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(4403. .4450)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(4545. .4785)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(4937. .5001)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(5102. .5191)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(5246. .5275)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(5437. .5484)
                     /note="AT_rich"
                     /rpt_type=tandem
```

```
     repeat_region   complement(5554. .5577)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(6302. .6350)
                     /note="AT_rich"

  Query Match             3.3%;  Score 70.4;  DB 3;  Length 14867;
  Best Local Similarity   43.9%;  Pred. No. 0.011;
  Matches  515;  Conservative    0;  Mismatches  631;  Indels   26;  Gaps    4;

Qy        893 TTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGT 952
Db       8215 TTAATTATTATTTATAAAAAAAATATATTAAATTATGATATATATATAATATTAATATAT 8274

Qy        953 TTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGT-----GAAAATGACGCCC 1007
Db       8275 ATTTAATTAATTAATTAAAATTAAAAATATATTAAATTATTTTTAATTAAATTATTATA 8334

Qy       1008 AAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTA 1067
Db       8335 TATTATTTATGAGTCTATTTATTTTATTTTTTATAATTAATTAATTAAAATTATATATTT 8394

Qy       1068 AATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACAT 1127
Db       8395 AATAATATATTTTGTTTCTTTGTATTAATATATTATTATTTTATTTATATTATTAAATTA 8454

Qy       1128 GATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGT 1187
Db       8455 AATTATATTAATTAATAAAAATTATAATATTATTTCATTTATTTTAATATATATTTATAT 8514

Qy       1188 TGTATGGATTAA-------ATTTTGAAGATATTTATATATTGAAGATGTTTATATATATT 1240
Db       8515 AATAATAATTATTTCCAATATAAATAAATAATGTATTTATTAAAGTTAATATAATTATTA 8574

Qy       1241 AGAGTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTT 1300
Db       8575 ATTTATTTTTATAAAATAAATATTGCTTTAATTCATTAATTATAATATTTATTATATTAT 8634

Qy       1301 GTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAAT 1360
Db       8635 TTTTAAAATAATTATGATATAATATTATATATATTAATAATTTAATTAGTATATAAAAAT 8694

Qy       1361 TTTTGTGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATAT 1420
Db       8695 TTAAACATATACATTATAATTT-----------ACACATAATTATTTATTTTATTTTAAT 8743

Qy       1421 TCCTCTTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGT 1480
Db       8744 TAATTTTTACTTATTTTAATTTAATTATTTTATTTTATATTATTTATTTATTATTTTAAT 8803

Qy       1481 CATTTGTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGA 1540
Db       8804 TAATTTATTTAATATATTAATTTATTTTTATTTTATTTATTTATTTTTATTTTATTTATT 8863

Qy       1541 GCTTCCAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAG 1600
Db       8864 TATTTATTTATTTTTATTTTATTTTATTTAATTATTTTAATTAATTTATTTAATATATTAA 8923

Qy       1601 TGTATCCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAG 1660
Db       8924 TTTATTTTTATTTTATTTATTTATTTAATATATTATTTATTTTTATTTTATTTATTTATT 8983

Qy       1661 ACCACCCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTC 1720
Db       8984 ATTTAAATTTATTTATTTAATATATTAATTTATTTTATTTTATTTATTTAATTATTTTAA 9043

Qy       1721 GAAGTTGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACA 1780
Db       9044 TTAATTTATTTAATATATTAATTTATTTTTATATTATTTATTTATTATTTAAATTTATTT 9103

Qy       1781 AACAAATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTC 1840
Db       9104 ATTTTATATTTTAATTAAATTATTTATTTTATATTATTTAATTATTATTTTAATTTATTT 9163

Qy       1841 ATTGTAAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTT 1900
Db       9164 ATTTTATATTATTTGTTTATTATTTTAGTTTATTTATTATTTTAATTAAATTATTTATTT 9223

Qy       1901 T---TGATATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTA 1957
Db       9224 TATATTATTTATATATTATTTTTTAATTTAATTTATTATATTAAATTTATTTAATATTTA 9283

Qy       1958 GTTTTGTCGTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAAT 2017
Db       9284 ATATTTTAATTTTATTTATTTAATTAATTAATTATTTATTTATTTTAATATAATTATTTT 9343

Qy       2018 TAGGAATCAAACATGATGATTAAGTTCACCAA 2049
Db       9344 TAATAAATATTTATTATATATTAATTTAAAAA 9375

RESULT 14
PFA929351
LOCUS       PFA929351             347050 bp    DNA     linear   INV 29-JAN-2003
DEFINITION  Plasmodium falciparum strain 3D7, chromosome 5, segment 1/4.
ACCESSION   AL929351 AL844504
VERSION     AL929351.1  GI:23504490
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., Mclean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
  MEDLINE   22255708
   PUBMED   12368867
REFERENCE   2  (bases 1 to 347050)
  AUTHORS   Devlin,K., Baker,S., Davies,P., Mungal,K., Berriman,M., Pain,A.,
            Hall,N., Bowman,S., Churcher,C., Quail,M. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .347050
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="5"
     repeat_region   1. .538
                     /note="Telomere"
     repeat_region   11583. .18389
                     /note="rep20"
     gene            join(20929. .26310,27215. .28456)
                     /gene="VAR"
                     /note="synonym: PFE0005w"
     CDS             join(20929. .26310,27215. .28456)
```

/gene="VAR"
/note="Similar to Plasmodium falciparum variant-specific surface protein var-2 TR:Q26033 (EMBL:L40609) (2664 aa) fasta scores: E(): 2.1e-94, 38.066% id in 2782 aa"
/codon_start=1
/product="erythrocyte membrane protein 1 (PfEMP1)"
/protein_id="CAD51367.1"
/db_xref="GI:23504491"
/db_xref="GOA:Q8I495"
/db_xref="SPTREMBL:Q8I495"
/translation="MGPPGITGTQGETAKHMFDRIGKQVYETVKNEAENYISELEGKL
SQATLLGERVSSLKTCQLVEDYRSKANGDVKRYPCANRSPVRFSDESRSQCTYNRIKD
NETDDNACGACAPYRRLHLCDYNLEKMGKTSTTKHDLLAEVCMAAKYEGDSIKTHYTI
HKHTNNDSAAELCTELARSFADIGDIIRGKDLYLGDIKKKQNGKKTEREKLEENLKRI
FGKIHEDVTNGKKEVLKTRYKDINDPEFFKLREDWWTANRATVWKAITCHAGESDKYF
RNTCNDSEHSGTFSQPNKYCRCNGDKPGEDKANVDPPTYFDYVPQYLRWFEEWAEDFC
RIRKHKLKNAKEQCREKYKSGTDRYCSRNGYDCTQTIRGRNILVSDSECTNCSVVCTP
FVKWIENKKLEFEKQKGKYTKEIEKANGTSNGTTIRTQYGTINNMYRKDFYQQLQSGY
GDVNAFLELLNKETTCKDHPKVEEKSDIDFNEGTEKTFSHTEYCETCPWCATKKKGID
GNWEEQKYEEGCENYLMKPIDESKSTDIDLLVKDTSGTTMVEKLGGLCNDSSKRTVQM
ENWQCHYEKKSQYEDGFDKDYCVLKDDKKKKPEHRTIKSYYTLFPNWINEMLKDSIDW
SKELKTCINNEKPTNCIRECKSKCDCFKKWVVQKEQEWKQLEEHYEKENFSGDFGPRI
SPYVTLEGNLQYSYLEMIRKYYAQEKPVQEIEQIIEKNKNNFEVKEDDNSITKFLQQE
KGIATKCIEKQEECKQQKKQQRQKQPADKVVSRSGASPDTASPDTKATEEKEEEEEEE
EEEDLGEESEEPEDQAVVDQDGQGETTEKKVPATTEEGSPKETTTPEKSVDVCKTVAE
LFSNVDNLKEACTQKYGGNNSRLGWKCIPTSGGEKATGGSGESTGSDATTGGSICVPP
RRRRLYVTPLTTWASGGTTQVETQASGGNTETSQVSGETTPQGQTPSESEAQTASQDP
SEKLRTAFIESAAVETFFLWDRYKKENTKRQSVLPLLEPINGDTISDDNNPEKLLKVG
KIPPDFLRLMFYTLGDYRDILFSSSNTSDTTGKETPSSSNDNLKNIVIEAGGTDEKDK
QKMKEIQQKIDKILKQSGTTPPTPVTHSPSSGTTPSSWWENNAKHIWKGMVCALTYKD
NTNGGPPTQDNDLKKALWDDKTNEPKKSEYKYDQVELKEENSGTGVTKGHAEAPGDNT
PLSKFVLRPPYFRYLEEWGETFCRERTRRLGKIKGECKVDDSGSGRGGVIKRQYSGDG
EDCDKVHEDPTTLPDLGSSCPKSCSSYRKWIKIKKDEYEKQQKAYEQQKKNCEKESEG
VENKSDICDQKFLQNLKNYGSIDSFLDRLKKGPCSKTYNESGEDNNKIFEDTEKTFGH
QNYCGPCSEFKINCEKGNCDKTKGQECNGKKTIDAQNFEQMGQTAKEFVMRVSDNDTT
KFEGHGLEDCNGAGIFKGIKENKWKCDKVCGYVVCKPKNGNGKNEGTYIIQIRALVRR
WVEYFLEDYNKIRKKLKPCINDGKEPKCIKTCDKKCNCAGKWIKLKQDEWENIKNRFN
EQYKNGNEGDYPVKTILEELISRIAAATDKAERGSLDKLKTSLGCNCSDNSQKKNTDE
RDIVECLLEKLQVKATSCQSQHSDKTEQQCQEYTPEPDEEDLLLEEENENQVENQKVG
NKAPAFCPPPEKPPKEEPDEKCGDKEEEKEQEEEKDKGNEEQSGAPSSPSPSEGTKEL
PPPPEPPEEAKPPEPVKPTPAPAPAAPPAAPAQPTLPADEPFNRDILATTIPFGIAFA
LGSIAFLFLKKKTKSTIDLLRVINIPKSDYDIPTKLSPNRYIPYTSGKYRGKRYIYLE
GDSGTDSGYTDHYSDITSSSESEYEEMDINDIYVPGSPKYKTLIEVVLEPSGNNTTAS
GNNTTASDTQNDIQNDGIPSNKFSDNEWNTLKDDFISQYLQSEQPNDLPNDYKSGNVT
LNTEPNILYFDKPEEKPFITSIHDRNLYTGEEYSYNINMSTNSMDDIPINRDNNVYSG
IDLINDALNGDYDIYDEMLKRKENELFGTNHVKQTSIHSVAKNTNSDPIHNQLELFHK
WLDRHRDMCEQWENHHERLAKLKEEWENETHSGNTHPSDSNKTLNTDVSIQIDMDDGK
PKKEFTNMDTILEDLDKYNEPYYDVQDDIYYDVNDHDTSTVNPNNMEKPSKVKIELDV
NKKTIKEKYPISDVWDI"

gene complement(29233. .29304)
/gene="PFE0010c"
/note="synonym: pseudo-VAR fragment"
/pseudo

CDS complement(29233. .29304)
/gene="PFE0010c"
/note="VAR fragment"
/pseudo
/codon_start=1

gene complement(29923. .30963)
/gene="RIF, degenerate"
/note="synonym: PFE0015c"
/pseudo

CDS complement(29923. .30963)
/gene="RIF, degenerate"
/note="Similar to Plasmodium falciparum rifin pfb1040W TR:O96293 (EMBL:AE001433) (345 aa) fasta scores: E(): 1.1e-50, 53.161% id in 348 aa. Contains nonsense mutations at apparent start position for exon 2"
/pseudo
/codon_start=1

misc_feature complement(30904. .30909)
/gene="RIF, degenerate"
/note="Possible splice donor"
/pseudo

gene complement(join(33056. .34126,34325. .34378))
/gene="RIF"
/note="synonym: PFE0020c"

CDS complement(join(33056. .34126,34325. .34378))
/gene="RIF"
/note="Similar to Plasmodium falciparum rifin pfb0030C TR:O96112 (EMBL:AE001367) (370 aa) fasta scores: E(): 3.3e-62, 53.125% id in 384 aa
1 probable transmembrane helix predicted for by TMHMM2.0 at aa 332-354"
/codon_start=1
/product="RIFIN"
/protein_id="CAD51370.1"
/db_xref="GI:23504492"
/db_xref="SPTREMBL:Q8I494"
/translation="MKFSYFNILLFSIPLNILINDHSKYSSCKHTSNSKTTKPHRSLY
ECGLYSPANNDNDPEMKRVMQQFEDRTSQRFHEYDERMQSKRMQCKEQCDKEIQKIIL
KDKLEKQMEQQLNTLETKIDTDDIPTCVCEKSLADKVEKGCLRCGYGLGTVAPTVGLI
GAVAVHVWKPMALEAAIEAAIAKSAAEISAAANAAGIQAGKIAVIESLKKLYVDYFWP
EMSNYILNMSHYNGVANLTAFIHEPKFNVCKDAGEVILDKCNAFDMGFGILKKDGVTN
GLLPKDAVPRVLKGIVGQAEGPAKVAADAARQTVTAEITEKETAAINTIFMSKQTAII
ASVVAIVVIVLIMIIIYLILRYRRKKKMKKKLQYIKLLKE"

gene complement(join(36464. .37324,37496. .37576))
/gene="RIF"
/note="synonym: PFE0025c"

CDS complement(join(36464. .37324,37496. .37576))
/gene="RIF"
/note="Similar to Plasmodium falciparum rifin pfb1005W TR:O96288 (EMBL:AE001432) (316 aa) fasta scores: E(): 8.3e-52, 54.574% id in 317 aa
2 probable transmembrane helices predicted for by TMHMM2.0 at aa 122-144 and 271-293"
/codon_start=1
/product="RIFIN"
/protein_id="CAD51371.1"
/db_xref="GI:23504493"
/db_xref="SPTREMBL:Q8I493"
/translation="MAINMKLHCSKILLFSLQLNILITSSYVNNKNKPYITPHTSAIT
SRVLNECDLYMPNYDKDADMKSVKENFDRQSSRRFEEYEERMITNRQKYKEQCEKDIQ
KIILKDKIEKLLEEKVEKCCLICGCGLGGGVAPFVGLFGGLAVNEMKKAAVVAATDVG
IKEAIKEVGSIFYLADGSTMEWTKMINAGNYSQKMSLVEIVTILKNKCEEDEALAGPL
FCKASEAATESGEVFEFSGNISRMAANAAEAAGNAANGKYAEMTSVGTICSNPVVISA
IVVVIIAVILLIIYLILRYRRKKKMKQKQQYTKLLNQ"

gene complement(39443. .40488)
/gene="STEVOR, degenerate"
/note="synonym: PFE0030c"
/pseudo

CDS complement(39443. .40488)
/gene="STEVOR, degenerate"
/note="Apparent single frameshift in exon II"
/pseudo
/codon_start=1

gene complement(41129. .41403)
/gene="RIF, degenerate"
/note="synonym: PFE0035c"
/pseudo

CDS complement(41129. .41403)
/gene="RIF, degenerate"
/note="RIF fragment"
/pseudo
/codon_start=1

gene complement(join(43327. .47481,47612. .47761))
/gene="MESA"
/note="synonym: PFE0040c"

CDS complement(join(43327. .47481,47612. .47761))
/gene="MESA"
/note="Similar to Plasmodium falciparum mature parasite-infected erythrocyte surface antigen TR:Q06166 (EMBL:AF056936) (1661 aa) fasta scores: E(): 3.6e-144, 81.331% id in 1698 aa
Identical to pfemp2 (EMBL:AF056936) except in internal repeat structures"

/codon_start=1

Query Match 3.3%; Score 70.4; DB 3; Length 347050;
Best Local Similarity 45.0%; Pred. No. 0.0071;
Matches 380; Conservative 0; Mismatches 461; Indels 3; Gaps 3;

```
Qy       1009 AAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAA 1068
Db     105990 ATATATATATACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTGTTAGAGAGAAAGGAA 106049

Qy       1069 ATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATG 1128
Db     106050 AAATTTCATATTCAGACAAAAATATATATAATTATATCATAACAATTTTAAAGAATCCTA 106109

Qy       1129 ATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTT 1188
Db     106110 TAATATAATAAGTATAGTTAGTCATATATATATATATATACGTAAATTGC-TGGGGAATT 106168

Qy       1189 GTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTA 1248
Db     106169 TTATGTCAAATAAATTGGCATATTCATTAAAATTAAATTCATTACATATATAAATATATT 106228

Qy       1249 TATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTA 1308
Db     106229 TATAACATTTTCTTATATTGTATAT-ATATATATATATATTTTTTTTTTTATTTATTT 106287

Qy       1309 GTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGT 1368
Db     106288 ATTTATTCTTTTCATTGAAATAATAATGAAAACTAAATTAAATTTATGAAGTTTTTATTA 106347

Qy       1369 AATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTT 1428
Db     106348 TATTTTTTTAATTTTATATGAAATTGTTATTTTATAATTTCCTATATTTTTTTTCAGATG 106407

Qy       1429 GCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTT 1488
Db     106408 ATGTTATCAATTCTGCTCTAATTGTTTTTAAAATTTATTCATTTTATATGTGTATTTAAT 106467

Qy       1489 GTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAA 1548
Db     106468 TATCTAATTATTTTATATATTTTTTTTTTAATTTATTTGTTAATTTTTTAAAATTTTAAA 106527

Qy       1549 GATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCA 1608
Db     106528 TTATTTATTATATAATATATATAT-ATATATATATATATATATATATATATATATATA 106586

Qy       1609 TATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCC 1668
Db     106587 TTTATATTTATATTTATATATATTTTTTTTTTTTCTTTCACGTTTGCATAATAATAGCCT 106646

Qy       1669 CTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGC 1728
Db     106647 TTTTATTAACTTTTTAAGATATATAATAAGAAAATATTAAATAAATAACCAAAAAATAAA 106706

Qy       1729 ACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATT 1788
Db     106707 ATATAAATAATCCATTTAATTAAATAAACAAAAAATAAAATATAAATAATCCATTTAATT 106766

Qy       1789 CAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAA 1848
Db     106767 AAATAAACAAATAAAATATCATATATATATATATATATGTGTGTGTGTGTGTTTCTTACA 106826

Qy       1849 TATT 1852
Db     106827 AGTT 106830
```

RESULT 15
ATT32A17/c
LOCUS ATT32A17 104386 bp DNA linear PLN 22-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T32A17 (ESSA project).
ACCESSION AL161813 AC007635

VERSION AL161813.1 GI:7321058
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 96453)
AUTHORS McCombie,R.W., Spiegel,L.A., Huang,E.N., Nascimento,L.U., de la Bastide,M., Vil,D.M., Preston,R.R., Matero,A., Shah,R., O'Shaughnessy,A., Rodriguez,M., Shekher,M., Schutz,K., See,L.H., Swaby,I., Habermann,K., Dedhia,N.N., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 95732 to 104386)
AUTHORS McCombie,R.W., Robben,J., Grymonprez,B., Bastiaens,I., Volckaert,G., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 104386)
AUTHORS EU Arabidopsis sequencing,project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
COMMENT On Mar 23, 2001 this sequence version replaced gi:5103797.
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.

FEATURES Location/Qualifiers

source 1. .104386
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"

misc_feature 1. .22908
/note="overlap to BAC T3F12; please refer to EMBL:AC002983 or analysis and annotation"

gene 22865. .24293
/gene="AT4g08700"

gene complement(join(22865. .23096,23227. .24293))
/gene="AT4g08700"

CDS complement(join(22865. .23096,23227. .24293))
/gene="AT4g08700"
/note="similarity to various hypothetical proteins of Arabidopsisthaliana
Contains Prokaryotic membrane lipoprotein lipid attachment site AA82-92;Prokaryotic membrane lipoprotein lipid attachment site AA225-235
contains EST gb:AI997220.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB82106.1"
/db_xref="GI:7321059"
/db_xref="SPTREMBL:Q9LDT1"
/translation="MLLLKEEDEGTRRTSVPTQLMKLKRTHWWILVFISIFFLISAQA
IAVLLGRFYYNEGGNSKWISTLVQTCGFPILYLPLCFLPASHSSSSSCSFKTLVWIYL
SLGFAIGLDNLLYSFGLLYLSASTYSILCSSQLAFNGVFSYYINSQKITCLILFSVLF
LSVSAVLVSLDDDSNSPSGDSKWSYLIGCLCTVFASLIYSLQLSLMQFSFENVLKSET
FSMVLEMQIYTSLVASCVAVIGLFASGEWMLLSVEMEEFHEGQVIYVLTLVGTAVSWQ
LGSVGAVALIFLVSSLFSNLIGTLSLIVTPLAAIAVFHDKLTEVKMVAMLIAFMGFGF
YIYQNYLDDLKVQRAREAQAEYLFLRAKVLCRIYIRNRSTYEVKYNGGFIENLGDTHV
VSEDHDVADLETNMVESNKHVDDSTLKASSSKRKMVIKGF"

exon complement(22865. .23096)
/gene="AT4g08700"
/number=1

intron complement(23097. .23226)
/gene="AT4g08700"
/number=1

exon complement(23227. .24293)
/gene="AT4g08700"
/number=2

gene 24970. .27800

```
                     /gene="AT4g08710"
     CDS             join(24970. .25451,25915. .26692,26786. .26947,27034. .27477,
                     27519. .27800)
                     /gene="AT4g08710"
                     /note="similarity to hypothetical proteins of Arabidopsis
                     thaliana Chromosome II and IV
                     contains EST gb:T04634, T20404, T20458, T46009, Z17562"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB82107.1"
                     /db_xref="GI:7321060"
                     /db_xref="SPTREMBL:Q9LD59"
                     /translation="MPPLKKRKRGVGTTVKSTTEETATATKETAPATKETAPATKETA
                     PTITKETAPTKETAPATKETAPTRTEEPSLTEQDPENVEEEESEEEEKEEEEKEEEEE
                     EEGEEEEEEEEEEEEEKEEEENVGGEESSDDSTRSLGKRPPPMRIWMRRHQLQWKFRRL
                     CNMNECLLQEIDEETEAVPPLSMYFPPSEYVKKIKLSTRCYIHELLTTFDKLEPPMSK
                     SERLWFENHPSFQHIFHMPRDPNHRLMGMWMLLLRTARIERKKEAWFIVNGVPIRYGI
                     LEHALISDFNCKNYKLGYKNTGNLDFKRKHFKDTVVKREDVREKLIGMVPEGERSKER
                     LRMMVLYFLSSIIIAPIKTGDKAPQVDEFFLKVMSDLTFCRNFQWGRYSFDYMLGTIS
                     HTVNHFNGSVTNNEKYIWPVPGFCLPMELLAFEAISQLREKFIEEIAEADPGCPRMCK
                     VRFKKNHLKGFPLDTIYAELGTTQVIDSIVTLGIEEQKLMKAITEEEDDVDVADMVVE
                     SWEKRRKEGRLIRFDELYQQDIQALQIIAGPKIPGAEDIPAVRGQELPSLVSVVQMIE
                     NLDKKMVDQMNKIIIMISDLDRRVESLEAFKDEQKAKEEKVHIDNCVILSNTMITRNQ
                     DEMNQEEAGDSREKDQEEDVVQEKDGDQEEDVVQEKDGDQEEDGVSKEKEQEKEKDPK
                     EKEKEKDPKEKDGGPSKPVEANVTRRITRRTKN"
     exon            24970. .25451
                     /gene="AT4g08710"
                     /number=1
     intron          25452. .25914
                     /gene="AT4g08710"
                     /number=1
     exon            25915. .26692
                     /gene="AT4g08710"
                     /number=2
     intron          26693. .26785
                     /gene="AT4g08710"
                     /number=2
     exon            26786. .26947
                     /gene="AT4g08710"
                     /number=3
     intron          26948. .27033
                     /gene="AT4g08710"
                     /number=3
     exon            27034. .27477
                     /gene="AT4g08710"
                     /number=4
     intron          27478. .27518
                     /gene="AT4g08710"
                     /number=4
     exon            27519. .27800
                     /gene="AT4g08710"
                     /number=5
     gene            28303. .31953
                     /gene="AT4g08720"
     gene            complement(join(28303. .28494,28565. .28761,29626. .30162,
                     30333. .31953))
                     /gene="AT4g08720"
     CDS             complement(join(28303. .28494,28565. .28761,29626. .30162,
                     30333. .31953))
                     /gene="AT4g08720"
                     /note="similarity to hypothetical protein F26B6.15 -
                     Arabidopsis thaliana, PID:g3242713"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB82108.1"
                     /db_xref="GI:7321061"
                     /db_xref="SPTREMBL:Q9LDY3"
                     /translation="MVEKVIIYFRWKDRAYSLMMNGKITLSMLESRICSKLGLDESKV
                     KLQMKYNAVMLGLTQEIHISDDEDVGVYISSGEMVSRYPLTVDVITPPELPEELSRVE
                     VVNKSSIGKNYDGLGSNEDEMRDNATIVIGELDGQGTELNGAIVVVDDTQDHEIGSRD
                     DRDDDDRGDEYVEPPPDVESSQFKKEWEDGIGLTLRQEFPSRAALQEVVDKGAFANSF
                     DYVIKKSDKERYVLTCSKANCDWRIQASCVQDTSIYSIRRYNKMHSCTRISKSSTRMR
                     KRKGTPELVAALLHDTFPGLLETPPPKVIMELVQTKLGVKVSYSTALRGKRQAVCDLK
                     GSAEDSYKDINCYLYMLKKVNDGTVTYMKLDESGKFQYLFIALGASIEGFQAMRKVII
                     VDATHLKNGYGGVLVFASARDPNRHHYIIAVGVLDGENDASWGWFFEKLLSVVPDTPE
                     LVFMSDRNSSLIKGIRNAYTAAHHGYCVWHLSQNVKAHSTNINRDVLAWRFMELSRIY
                     TWSEFEREYRVFKVRYPTAVNYLEDSTVKEKWGSVDNKFVPLVENYLHNLWVDAEKLK
                     VTELNTFRLEYNVRGVEGNEYLVNLHFKTCSCKVFDIQKYPCIHALAAFIVFDSDTTR
                     TRGMELHELVSKYYWAELWALAYYGTIYLVPGRSQWDIPDDIKALKVLPLPRKKKKGR
                     TKVLRIPSTGERRPKRQRTQNKRRPRQSLQWLLFGTRTMDRLCESDPWYGEMKSAKRI
                     MQQLEEVAMMEGIPPICPCGGRILDIISEKDGDKGKRFYECTNYKNDGLHIRKLWDKA
                     MVEEVNRLREQVDNHHQKIQSLEYYNQQMRSKLDEIQKKMGTLWRICCGE"
     exon            complement(28303. .28494)
                     /gene="AT4g08720"
                     /number=1
     intron          complement(28495. .28564)
                     /gene="AT4g08720"
                     /number=1
     exon            complement(28565. .28761)
                     /gene="AT4g08720"
                     /number=2
     intron          complement(28762. .29625)
                     /gene="AT4g08720"
                     /number=2
     exon            complement(29626. .30162)
                     /gene="AT4g08720"
                     /number=3
     intron          complement(30163. .30332)
                     /gene="AT4g08720"
                     /number=3
     exon            complement(30333. .31953)
                     /gene="AT4g08720"
                     /number=4
     gene            34618. .36007
                     /gene="AT4g08730"
     gene            complement(join(34618. .34852,35539. .35690,35828. .36007))
                     /gene="AT4g08730"
     CDS             complement(join(34618. .34852,35539. .35690,35828. .36007))
                     /gene="AT4g08730"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB82109.1"
                     /db_xref="GI:7321062"
                     /db_xref="SPTREMBL:Q9LDC3"
                     /translation="MEGFTNFLQVMETKVDLLARKVDLLTSKMDLCATNVDLLTSKMN
                     LCATNSDLKAMLTSMKGKNVPAKKTLAMAASDSEDSSDSSSSDDIELCFRFLIKIVIQ
                     LRPSYFIIIFAYIDLKEGANKALELNGSDMAGKELVVKTALLMRDSYLGYSCNGGIGG
                     RFGGRFGSVGGIFGGGRCGGLGQYYAQV"
     exon            complement(34618. .34852)
                     /gene="AT4g08730"
                     /number=1
     intron          complement(34853. .35538)
                     /gene="AT4g08730"
```

```
  Query Match             3.2%;  Score 69.2;  DB 8;  Length 104386;
  Best Local Similarity   85.6%;  Pred. No. 0.013;
  Matches   77;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy         88 GGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAAT 147
              || |||| ||||| ||||||| ||||||||||||||| ||  ||||||||||| |||||
Db      98373 GGTATGGGACCATATTTTTAGAATTTACTTTGTAATTTCTATATCTTTCATTTTTTGAAA 98314

Qy        148 TGATATTTACATTTTTATCAAAAAAAAAGT 177
              |||||||||||| |||| | ||||| |||
Db      98313 TGATATTTACATTCTTATAAGAAAAATAGT 98284


Search completed: July 31, 2004, 06:47:47
Job time : 8378 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 02:09:18 ; Search time 859 Seconds
(without alignments)
10627.913 Million cell updates/sec

Title: US-10-027-880-1
Perfect score: 2149
Sequence: 1 gtcgaattgtgatatattgt.....actaattgctaattaatcag 2149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2149	100.0	2149	4	AAF25316	Aaf25316 Nucleotid
	2	2149	100.0	4280	4	AAF25317	Aaf25317 Nucleotid
	3	2149	100.0	4309	4	AAF25319	Aaf25319 Nucleotid
	4	2149	100.0	4413	4	AAF25318	Aaf25318 Nucleotid
c	5	63.6	3.0	7008	4	AAS46532	Aas46532 Tumour su
	6	60.2	2.8	6175	6	ABL33307	Abl33307 Human imm
	7	59.8	2.8	7597	6	ABL33013	Abl33013 Human imm
	8	59.6	2.8	5822	6	ABL33097	Abl33097 Human imm
	9	59.4	2.8	8920	2	AAQ62924	Aaq62924 Carbamoyl
	10	59.4	2.8	33053	6	ABQ67005	Abq67005 Human ang
	11	58.6	2.7	5413	4	AAS46694	Aas46694 Tumour su
	12	58.2	2.7	6045	6	ABK31541	Abk31541 Signal tr
	13	58.2	2.7	6045	6	ABL70624	Abl70624 Chemicall
	14	58.2	2.7	8056	7	ABZ10100	Abz10100 Haematopo
	15	57.6	2.7	37515	6	ABQ66998	Abq66998 Human ang
	16	57.2	2.7	8056	7	ABZ10246	Abz10246 Haematopo
	17	56.8	2.6	8277	6	ABL33328	Abl33328 Human imm
	18	56.6	2.6	6292	4	AAS46735	Aas46735 Tumour su
c	19	56.6	2.6	8056	7	ABZ10246	Abz10246 Haematopo
	20	56.4	2.6	15951	6	ABL33680	Abl33680 Human imm
	21	56.4	2.6	15951	6	ABL34580	Abl34580 Human met
	22	56.4	2.6	15951	6	ABL70373	Abl70373 Chemicall
c	23	56.2	2.6	8056	7	ABZ10100	Abz10100 Haematopo
c	24	56.2	2.6	17280	4	AAS46772	Aas46772 Tumour su
	25	56	2.6	37973	6	ABL34197	Abl34197 Human imm
	26	55.8	2.6	6076	4	AAS46664	Aas46664 Tumour su
c	27	55.6	2.6	393	7	ABX39417	Abx39417 Bovine ES
	28	55.6	2.6	7072	6	ABK31471	Abk31471 Signal tr
	29	55.6	2.6	7072	6	ABL70566	Abl70566 Chemicall
	30	55.6	2.6	7072	6	AAS61385	Aas61385 Human gen
	31	55.6	2.6	7922	6	ABN80086	Abn80086 Human che
	32	55.4	2.6	12237	6	ABL34358	Abl34358 Human imm
	33	55.4	2.6	15224	6	ABK31358	Abk31358 Signal tr
	34	55.4	2.6	15224	6	ABL70317	Abl70317 Chemicall
	35	55.4	2.6	15224	6	AAS61262	Aas61262 Human gen
	36	55.4	2.6	110000	7	AAD53224_2	Continuation (3 of
	37	55	2.6	11745	6	ABK28332	Abk28332 DNA trans
c	38	55	2.6	113515	6	ABL34174	Abl34174 Human imm
	39	54.8	2.6	5297	4	AAS46513	Aas46513 Tumour su
	40	54.8	2.6	14919	4	AAS46506	Aas46506 Tumour su
	41	54.6	2.5	8244	4	AAS46395	Aas46395 Tumour su
	42	54.6	2.5	8244	6	ABL32992	Abl32992 Human imm
	43	54.6	2.5	8244	6	ABQ67031	Abq67031 Human ang
	44	54.6	2.5	8244	9	ADB54133	Adb54133 Pretreate
	45	54.6	2.5	13131	6	ABL92248	Abl92248 Chemicall

ALIGNMENTS

RESULT 1
AAF25316
ID AAF25316 standard; DNA; 2149 BP.
XX
AC AAF25316;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plant promoter for expression in roots.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR001768.
XX
PR 25-JUN-1999; 99FR-00008185.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful for
PT controlling expression of pest or drought resistance genes, and related
PT transgenic plants.
XX
PS Claim 1; Fig 1; 92pp; French.
XX
CC The present sequence represents a plant promoter that directs expression
CC of a selected sequence in root cells at all stages of development of a
CC plant. The plant promoter is used to control expression of genes in a
CC root-specific manner, especially genes that provide resistance to
CC parasites, pests (nematodes or fungi), water and salt stress, or alter
CC sugar content or nitrogen transport. Fragments of the promoter are useful
CC as probes or primers to detect or amplify at least part of the promoter
XX
SQ Sequence 2149 BP; 693 A; 365 C; 335 G; 756 T; 0 U; 0 Other;

Query Match 100.0%; Score 2149; DB 4; Length 2149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
Db       1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy      61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
Db      61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy     121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
Db     121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy     181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
Db     181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy     241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
Db     241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy     301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
Db     301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy     361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
Db     361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy     421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
Db     421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy     481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
Db     481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy     541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
Db     541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy     601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
Db     601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy     661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
Db     661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy     721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
Db     721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy     781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
Db     781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy     841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
Db     841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy     901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
Db     901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy     961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
Db     961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy    1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
Db    1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy    1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
Db    1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy    1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
Db    1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy    1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
Db    1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy    1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
Db    1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy    1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
Db    1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy    1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
Db    1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy    1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
Db    1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy    1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
Db    1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy    1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
Db    1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy    1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
Db    1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy    1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
Db    1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy    1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
Db    1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy    1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
Db    1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy    1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
Db    1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy    1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
Db    1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy    1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
Db    1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy    2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
Db    2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy    2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

```
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149


RESULT 2
AAF25317
ID   AAF25317 standard; DNA; 4280 BP.
XX
AC   AAF25317;
XX
DT   30-APR-2001  (first entry)
XX
DE   Nucleotide sequence of a root promoter and the GSU gene coding region.
XX
KW   Plant promoter; root cell; root-specific expression; parasite resistance;
KW   nematode resistance; fungal resistance; water stress; salt stress;
KW   sugar content; nitrogen transport; GUS gene; ss.
XX
OS   Synthetic.
OS   Arabidopsis thaliana.
OS   Unidentified.
XX
PN   WO200100833-A1.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-FR001768.
XX
PR   25-JUN-1999;   99FR-00008185.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Hoffmann B,  Mollier P,  Pelletier G;
XX
DR   WPI; 2001-102893/11.
XX
PT   New constitutive plant promoter active specifically in roots, useful for
PT   controlling expression of pest or drought resistance genes, and related
PT   transgenic plants.
XX
PS   Claim 5; Page 69-71; 92pp; French.
XX
CC   The present sequence represents a plant promoter linked to a GUS coding
CC   region. The plant promoter directs expression of a selected sequence in
CC   root cells at all stages of development of a plant. The plant promoter is
CC   used to control expression of genes in a root-specific manner, especially
CC   genes that provide resistance to parasites, pests (nematodes or fungi),
CC   water and salt stress, or alter sugar content or nitrogen transport.
CC   Fragments of the promoter are useful as probes or primers to detect or
CC   amplify at least part of the promoter
XX
SQ   Sequence 4280 BP; 1256 A; 859 C; 907 G; 1258 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 2149;  DB 4;  Length 4280;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

RESULT 3
AAF25319
ID AAF25319 standard; DNA; 4309 BP.
XX
AC AAF25319;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plant promoter insert in plasmid pBin19.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Synthetic.
OS Arabidopsis thaliana.
OS Unidentified.
XX

```
FH   Key             Location/Qualifiers
FT   promoter        30. .2178
FT                   /*tag= a
FT   misc_feature    2179. .4309
FT                   /*tag= b
FT                   /note= "GUS coding region"
XX
PN   WO200100833-A1.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-FR001768.
XX
PR   25-JUN-1999;   99FR-00008185.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Hoffmann B,  Mollier P,  Pelletier G;
XX
DR   WPI; 2001-102893/11.
XX
PT   New constitutive plant promoter active specifically in roots, useful for
PT   controlling expression of pest or drought resistance genes, and related
PT   transgenic plants.
XX
PS   Example 3; Page 73-75; 92pp; French.
XX
CC   The present sequence represents a plant promoter insert in plasmid
CC   pBin19. The plant promoter directs expression of a selected sequence in
CC   root cells at all stages of development of a plant. The plant promoter is
CC   used to control expression of genes in a root-specific manner, especially
CC   genes that provide resistance to parasites, pests (nematodes or fungi),
CC   water and salt stress, or alter sugar content or nitrogen transport.
CC   Fragments of the promoter are useful as probes or primers to detect or
CC   amplify at least part of the promoter
XX
SQ   Sequence 4309 BP; 1262 A; 868 C; 915 G; 1264 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2149;  DB 4;  Length 4309;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         30 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 89

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 149

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 209

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 269

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 329

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 389

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 449

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 509
```

```
Qy       481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
Db       510 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 569

Qy       541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
Db       570 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 629

Qy       601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
Db       630 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 689

Qy       661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
Db       690 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 749

Qy       721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
Db       750 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 809

Qy       781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
Db       810 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 869

Qy       841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
Db       870 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 929

Qy       901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
Db       930 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 989

Qy       961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
Db       990 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1049

Qy      1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
Db      1050 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1109

Qy      1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
Db      1110 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1169

Qy      1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
Db      1170 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1229

Qy      1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
Db      1230 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1289

Qy      1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
Db      1290 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1349

Qy      1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
Db      1350 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1409

Qy      1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
Db      1410 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1469

Qy      1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
Db      1470 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1529

Qy      1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
Db      1530 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1589

Qy      1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
Db      1590 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1649

Qy      1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
Db      1650 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1709

Qy      1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
Db      1710 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1769

Qy      1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
Db      1770 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1829

Qy      1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
Db      1830 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1889

Qy      1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
Db      1890 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1949

Qy      1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
Db      1950 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2009

Qy      1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
Db      2010 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2069

Qy      2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
Db      2070 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2129

Qy      2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
Db      2130 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2178
```

RESULT 4
AAF25318
ID AAF25318 standard; DNA; 4413 BP.
XX
AC AAF25318;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of an Arabidopsis thaliana genomic fragment.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR001768.
XX
PR 25-JUN-1999; 99FR-00008185.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful for
PT controlling expression of pest or drought resistance genes, and related

PT transgenic plants.
XX
PS Claim 3; Page 71-73; 92pp; French.
XX
CC The present sequence represents a genomic fragment of Arabidopsis
CC thaliana. The sequence contains a plant promoter that directs expression
CC of a selected sequence in root cells at all stages of development of a
CC plant. The plant promoter is used to control expression of genes in a
CC root-specific manner, especially genes that provide resistance to
CC parasites, pests (nematodes or fungi), water and salt stress, or alter
CC sugar content or nitrogen transport. Fragments of the promoter are useful
CC as probes or primers to detect or amplify at least part of the promoter
XX
SQ Sequence 4413 BP; 1427 A; 731 C; 702 G; 1553 T; 0 U; 0 Other;

Query Match 100.0%; Score 2149; DB 4; Length 4413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615

Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675

Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735

Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795

Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855

Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915

Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975

Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035

Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095

Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155

Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215

Qy       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAACATGATAGACATTTAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAACATGATAGACATTTAG 1275

Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335

Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395

Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455

Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515

Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575

Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635

Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695

Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755

Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815

Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875

Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935

Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995

Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
```

```
Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
Db       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115

Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
Db       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175

Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
Db       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235

Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
Db       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
```

```
RESULT 5
AAS46532/c
ID   AAS46532 standard; DNA; 7008 BP.
XX
AC   AAS46532;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #254.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 254; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7008 BP; 2141 A; 26 C; 1239 G; 3602 T; 0 U; 0 Other;

  Query Match             3.0%;  Score 63.6;  DB 4;  Length 7008;
  Best Local Similarity   44.1%;  Pred. No. 0.0077;
  Matches  264;  Conservative    0;  Mismatches  334;  Indels    0;  Gaps    0;

Qy        116 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 175
Db       4997 TTTCTCATATATAAACCTTCCTTAACAATAATTATTAAAAATTTCTTACTAATCACAAAA 4938

Qy        176 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 235
Db       4937 CTACTATTTAAATAATAATTTAAAATAAAAAAAACCAATATATTTCATCTCAAATCAAA 4878

Qy        236 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 295
Db       4877 TCAAATATAATCCAACCAACATCTATCAATACTAAAACTAAACTCCATAAAAACAAACAT 4818

Qy        296 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 355
Db       4817 CATATCCATTTCCCTAACACTCAACCAAAAATTTTTATATAATAAAAACTTAATAAATAT 4758

Qy        356 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 415
Db       4757 TTAAATATATTAATAATTAAATAAATAAAAAACCGAATAATAAACCAATAAATAACTACT 4698

Qy        416 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 475
Db       4697 CAACCGAAAACACCGAAAACACAAAAATAAACAAAACATAAACCTTAATCTTCAAAAATA 4638

Qy        476 TTATTCTTAATAACATACTCTATATTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 535
Db       4637 TATACATTTATCACTTAATTTATATTTTAACAAATATTTATTTAATATCAACTATTTTC 4578

Qy        536 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 595
Db       4577 CAAACACTTTATTAAATACTACAATAAATAAAACAACTATATTTACTCCCTAAAAATAAA 4518

Qy        596 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 655
Db       4517 AATATTACATAATAATCGTATAAATAACTATAAAATTAAAATTATTTAAATACAATAAA 4458

Qy        656 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAGAAAA 713
Db       4457 AAATTTTATAAAAATCGTACAATAAAAAAAAATAATCTAAATACAAAAAAAAATAAAA 4400
```

```
RESULT 6
ABL33307
ID   ABL33307 standard; DNA; 6175 BP.
XX
AC   ABL33307;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1280.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
```

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1280; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6175 BP; 1506 A; 154 C; 1357 G; 3158 T; 0 U; 0 Other;

Query Match 2.8%; Score 60.2; DB 6; Length 6175;
Best Local Similarity 50.2%; Pred. No. 0.035;
Matches 149; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 1113 TAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTG 1172
Db 5561 TATATGTTAATATTTTATATATGTTAATATTTTATATATTTATATTTATTTATATGTTTA 5620

Qy 1173 GATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTA 1232
Db 5621 TATTTTATATATGTTTATATTTTATATTTTATATATATGTTTATATTTTATATTTATTTA 5680

Qy 1233 TATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTAC 1292
Db 5681 TATATATTTGTTTATATATATTTGTATATATATATGTTTATATATATGTTTATATATTAT 5740

Qy 1293 TCTTTTTTGTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTT 1352
Db 5741 ATATATTTGTTTATATATATGTTTATATTTTATAGTTGTTTATTTATTATTTATACGTTT 5800

Qy 1353 AAGGAAATTTTTGTGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTAT 1409
Db 5801 ATATTAGTTTTTTTATATTTTATTGATATTTGTTTATTTATATTTTATATTATTTAT 5857

RESULT 7
ABL33013
ID ABL33013 standard; DNA; 7597 BP.
XX
AC ABL33013;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 986.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 986; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 U; 0 Other;

Query Match 2.8%; Score 59.8; DB 6; Length 7597;
Best Local Similarity 42.7%; Pred. No. 0.043;
Matches 429; Conservative 0; Mismatches 567; Indels 9; Gaps 2;

Qy 1025 TCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCA 1084
Db 110 TATTTTTTTATTTATGTATTTTTTTATTATTATTTATTTATTTATTTGTTTATGTATTTA 169

Qy 1085 CTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTTA 1144
Db 170 TTTGTTTATTTATTTATTTATTTATTGTATTTATTTATTCGTTTATTATTTATTTATTTAT 229

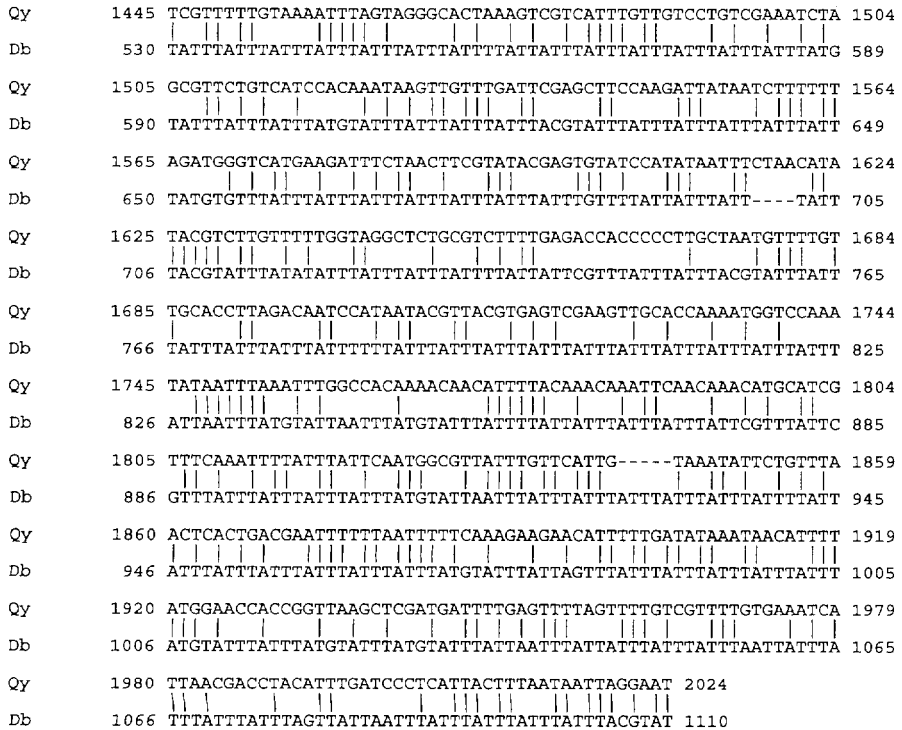
Qy 1145 AAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTT 1204
Db 230 TTATTTATGAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 289

Qy 1205 GAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATT 1264
Db 290 GTATTTATTTATTATTTATTTATTTTATTATTTATTTATTTATTTATTTATTTATTTATG 349

Qy 1265 GATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCATC 1324
Db 350 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAATTATTTATGTATTTATT 409

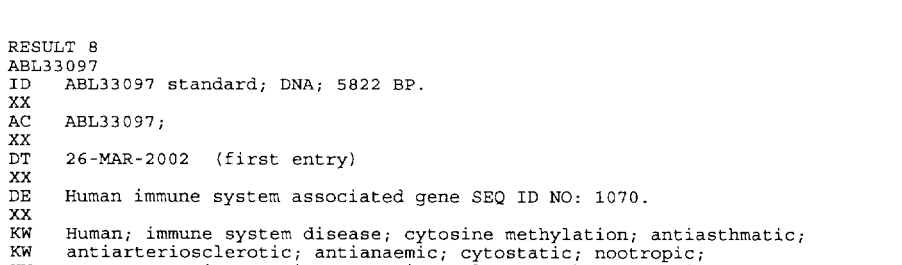
Qy 1325 CTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTTTT 1384
Db 410 TATTTATTTATTTATTTATTTATTTATTTATTTATTTTATTATTTATTTATTTATTTATT 469

Qy 1385 CGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAAAAGA 1444
Db 470 TATTTATGTATTTATTTATTTATGTATTTATTTATTTATGTATTTATTTATTTATTTATT 529



RESULT 8
ABL33097
ID ABL33097 standard; DNA; 5822 BP.
XX
AC ABL33097;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1070.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1070; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5822 BP; 1880 A; 30 C; 1198 G; 2714 T; 0 U; 0 Other;

Query Match 2.8%; Score 59.6; DB 6; Length 5822;
Best Local Similarity 51.1%; Pred. No. 0.045;
Matches 189; Conservative 0; Mismatches 179; Indels 2; Gaps 2;



RESULT 9
AAQ62924
ID AAQ62924 standard; cDNA; 8920 BP.
XX
AC AAQ62924;
XX
DT 25-MAR-2003 (revised)
DT 06-DEC-1994 (first entry)
XX
DE Carbamoyl-phosphate-synthetase II.
XX
KW Carbamoyl-phosphate-synthetase II; CPSII; psCPSII gene; malaria; ss.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1226. .8401
FT /*tag= a

```
FT                   /EC_number= "6.3.5.5"
XX
PN   WO9412643-A1.
XX
PD   09-JUN-1994.
XX
PF   02-DEC-1993;   93WO-AU000617.
XX
PR   03-DEC-1992;   92AU-00006206.
PR   16-DEC-1992;   92AU-00006380.
XX
PA   (UNIX ) UNISEARCH LTD.
XX
PI   Stewart TS,  Flores MV,  Osullivan WJ;
XX
DR   WPI; 1994-200271/24.
DR   P-PSDB; AAR55694.
XX
PT   Nucleic acid encoding carbamoyl phosphate synthetase II - isolated from
PT   Plasmodium falciparum, used to develop prods. for the treatment of
PT   malaria.
XX
PS   Disclosure; Page 6-16; 31pp; English.
XX
CC   The cDNA sequence encoding the carbamoyl-phosphate-transferase II (CPSII)
CC   of Plasmodium falciparum was determined. The cDNA encodes a protein that
CC   includes 2 insert sequences not found in other CPSII proteins. The first
CC   separates the putative structural subdomain and the glutaminase subdomain
CC   of the glutamine-amidotransferase subunit of CPSII, while the second
CC   separates 2 ATP binding subdomains of the CPSII subunit, CPSa and CPSb.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 8920 BP; 3836 A; 774 C; 1232 G; 3078 T; 0 U; 0 Other;

  Query Match             2.8%;  Score 59.4;  DB 2;  Length 8920;
  Best Local Similarity   45.7%;  Pred. No. 0.052;
  Matches  207;  Conservative    0;  Mismatches  246;  Indels    0;  Gaps    0;

Qy        266 ATATATACTGATATTGATATGCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATC 325
Db        333 ATATATATATATATGTTTGTGTGTTCATTTGTTTATAAAATTACTTGAAATATAAAACTT 392

Qy        326 AGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAA 385
Db        393 ATTAATATATTTCCAATTAATATGAATACAATTATTAATATTTTGATGTGTACACATTAA 452

Qy        386 AAAAGTGGTATACTGGTATATACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGG 445
Db        453 TATAGTTTTACACTTCTTATAATAAAACCATCCTATATATTATACACAATATATAATACT 512

Qy        446 AGAAAATGGACGTGTCTCTTTGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAA 505
Db        513 CCCCAATATTGTGGTTCCTATAATTTTATTTATATATTTATTTATTAATTTATTCATTTA 572

Qy        506 ACACTTCGATGTCTCGCTTAAATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTA 565
Db        573 TTTATTTTTTTTCTTAGTTTATAAAATAGTAATTCTACTAATTTAAAAAAAAAAAAAAAA 632

Qy        566 AAGAAAATTCGTCGAAGCCACAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCAT 625
Db        633 AAAAAAAAAAAAAGAAAAAAAAAAAAATTTACATATGAAAAATGAACTTGTATATGTAAAT 692

Qy        626 CAAAAATATATTAATAGAAAAGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAA 685
Db        693 TTATAAATATTTTAAACATAAATATAAATGTATAAAAAAAAAAAAGAAAAATGGGAAAAA 752

Qy        686 AAAAGATAAGAAGGTAGGCAGAAGAAAACGTAT 718
Db        753 ATAATATAGATATATATATAAATATATATATAT 785


RESULT 10
ABQ67005

ID   ABQ67005 standard; DNA; 33053 BP.
XX
AC   ABQ67005;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Human angiogenesis associated polynucleotide SEQ ID NO 35.
XX
KW   Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW   inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW   macular degeneration; inflammatory bowel disease; Crohn's disease;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiarteriosclerotic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200246454-A2.
XX
PD   13-JUN-2002.
XX
PF   06-DEC-2001; 2001WO-EP014320.
XX
PR   06-DEC-2000; 2000DE-01061338.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Schacht O;
XX
DR   WPI; 2002-500450/53.
XX
PT   New nucleic acid fragments from chemically treated angiogenesis-
PT   associated genes, useful for determining methylation status, e.g. in
PT   diagnosis or treatment of cancer.
XX
PS   Claim 1; SEQ ID NO 35; 41pp + Sequence Listing; German.
XX
CC   The invention relates to a nucleic acid (I) comprising a segment of 18
CC   bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC   having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC   related oligomers, are used to evaluate the methylation status and/or
CC   single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC   diagnosis and treatment of eye diseases, proliferative retinopathy,
CC   neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC   diabetic retinopathy, macular degeneration caused by neovascularisation,
CC   psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC   Crohn's disease. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 33053 BP; 9206 A; 283 C; 6714 G; 16850 T; 0 U; 0 Other;

  Query Match             2.8%;  Score 59.4;  DB 6;  Length 33053;
  Best Local Similarity   51.1%;  Pred. No. 0.064;
  Matches  166;  Conservative    0;  Mismatches  156;  Indels    3;  Gaps    1;

Qy       1064 GTTAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAA 1123
Db       8962 GTGAAAAATTTAATAATAGTTTTATAATATTAAAAATTATAGTTAAAATAATTTAAAAAT 9021

Qy       1124 ACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGA 1183
Db       9022 ATATAATTTATATTTTTGTAATTAGAAGAAATGTTTATTAAAAATATAAGAATATGTTAA 9081

Qy       1184 TTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGA 1243
Db       9082 TTTGATTTGAGTAAGTATTTTTGTGAAATTGTGGTTTTGG---TGTTTGTTTTTTTTTTA 9138

Qy       1244 GTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTT 1303
Db       9139 TATTAATTGATAAAAAATTTTATTTATTGTAAATTAATTTTGTTCGTCGGGTTTTTTATC 9198

Qy       1304 GCGTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTT 1363
```

```
Db       9199 GGTAATAATTTTATTTAATAGTAGTAAAAATAAGTTATTTGAAAGTGATAGCGATTATT 9258

Qy       1364 TGTGTAATTCATAGTCTTTTTCGTA 1388

Db       9259 TTAGTTTTTTGTATATTTTTGTTTA 9283
```

RESULT 11
AAS46694
ID   AAS46694 standard; DNA; 5413 BP.
XX
AC   AAS46694;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #417.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 417; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5413 BP; 1391 A; 65 C; 1102 G; 2855 T; 0 U; 0 Other;

Query Match 2.7%; Score 58.6; DB 4; Length 5413;
Best Local Similarity 44.0%; Pred. No. 0.069;
Matches 351; Conservative 0; Mismatches 434; Indels 12; Gaps 2;

```
Qy       1135 ATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGG 1194

Db       1249 ATGTATTTTAGATTTTTTATATTAAGATGAAATTTATTTTCGTAGTTTATTTTGTTATAT 1308

Qy       1195 ATTAAATTTTGAAGATATTATATATTGAAGATGTTTATATATAT-TAGAGTTTATATAG 1253

Db       1309 ATATATTTTTTATATATATATATTTATATATATTTTATATATATATATATTTTATATAT 1368

Qy       1254 CAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCT 1313

Db       1369 ATATATTTTTATATATATATATAGAGGTATTTTTTATTTTATTTTTATAGAGGTATTTAT 1428

Qy       1314 TTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTC 1373

Db       1429 ATATATATATTTTTATATAAATATATATATTTATTTTTTATAGGTATTTTTTTATATATT 1488

Qy       1374 A-----------TAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTC 1422

Db       1489 TATATATAGATTTATTTTTTTATATATAGGTATTTATATTTATATTTAGTAAATATATAT 1548

Qy       1423 CTCTTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCA 1482

Db       1549 ATTTATTTTTTATAAGTGTTTTTTTATTATATATATTTATTTTTTATATATAGGTATTTA 1608

Qy       1483 TTTGTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGC 1542

Db       1609 TTTTTATATACGTATTTATATATATATTTATGTATATATTTTATATATATTTTTTTTATA 1668

Qy       1543 TTCCAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTG 1602

Db       1669 TATATAGGTTTTTATATTTAGGAATATATATATTCGTTTATATATTTATATAGTTATATT 1728

Qy       1603 TATCCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGAC 1662

Db       1729 TATATATATATATATACGTATTTATTTTTATATATACGTATTTATTTTATATATATATAT 1788

Qy       1663 CACCCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGA 1722

Db       1789 TTTTTTTTTTTTAATATATATTTATATATATATATAGGTATTTTTATATATTTATTTATA 1848

Qy       1723 AGTTGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAA 1782

Db       1849 TTTATGTATATATTTATATATTTATTTTATATATATATATAGGTATTTTTTTTATAGATA 1908

Qy       1783 CAAATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCAT 1842

Db       1909 TTTTTATATATTTATATTTATATATATATAAATATTTATATATAGGTATTTTTATATATT 1968

Qy       1843 TGTAAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTT 1902

Db       1969 TTTATATATATAGTTTGTGTTTTGAATTATTTTTGTTTTTTTTAAATATTTTATGTTATT 2028

Qy       1903 GATATAAATAACATTTT 1919

Db       2029 TTTTTATTTAATGTTTT 2045
```

RESULT 12
ABK31541
ID   ABK31541 standard; DNA; 6045 BP.
XX
AC   ABK31541;
XX
DT   23-APR-2002 (first entry)
XX
DE   Signal transduction associated gene modified complementary DNA #192.
XX
KW   Human; signal transduction associated gene; cytosine methylation state;
KW   CpG island; signal transduction associated disease; solid tumour; cancer;
KW   antitumour; cytostatic; mutant; ds.
XX

OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 384; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 6045 BP; 2228 A; 31 C; 950 G; 2836 T; 0 U; 0 Other;

Query Match 2.7%; Score 58.2; DB 6; Length 6045;
Best Local Similarity 44.6%; Pred. No. 0.084;
Matches 283; Conservative 0; Mismatches 343; Indels 9; Gaps 1;

Qy 856 TATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGTGATATATAATCTAAC 915
Db 3012 TAAGATATATTAAAATTTTTGATATTTTTTTGAAATATTATTTATTAATTATTATTTATA 3071

Qy 916 TTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTGACCTCAGTCCCTAGT 975
Db 3072 TGAAAGGTTATTTAAAGTATTTTTGAGTTATATAATTTTTTGTTGATTTATATAGTAATT 3131

Qy 976 TAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGCAA 1035
Db 3132 AAATTATTATATGTTAAGTAATAATAAATTTTAGAAATTTTTATTTTGGATTATTTTAAA 3191

Qy 1036 CTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCGTATTA 1095
Db 3192 TAAGGATTTTAGTGATTATTTTATATATTTTTTTTTTAGGTTAGTTATTTTTTAAAAATA 3251

Qy 1096 AATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAAT 1155
Db 3252 AAAATGTTAAT---------ATAAGGTTAGAGTTTTTATATGATAATAAGTAGATAATTT 3302

Qy 1156 GTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTA 1215
Db 3303 TATATATTAAATGTGTTTTGTAGATAAGTTATAGATACGATTTTTTTTTTAAAGAAGTTT 3362

Qy 1216 TATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGT 1275
Db 3363 AAAAATTTAGATAGTTATATTTTTATTTTTTTGAGAAAAATATAGAGTATATTTTTTATT 3422

Qy 1276 TGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAG 1335
Db 3423 ATTTGTTTTATAAATGGAGTTATTTATAAATTATGGAGTAATAATTAAATGTTAATGATA 3482

Qy 1336 AAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTTTTCGTAACCACAG 1395
Db 3483 AATATTTTTTTTTATTATAGGAATTTTTTATGATTTTTAAATTTTTTTATAGTATGTTTTA 3542

Qy 1396 TTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAAAAGATCGTTTTTGTA 1455
Db 3543 GTTTAGTTTGTTAAAGGTAATGTGAAGGTTTACGAATAATTTAGGAATCGAGTTTATATG 3602

Qy 1456 AAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGT 1490
Db 3603 GTTTTTTGAAGTTAAATTTATAAATTATTTTTTT 3637

RESULT 13
ABL70624
ID ABL70624 standard; DNA; 6045 BP.
XX
AC ABL70624;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence complementary to#257.
XX
KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007471.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX
PS Claim 1; SEQ ID NO 514; 24pp + Sequence Listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the

```
CC   European Patent Office
XX
SQ   Sequence 6045 BP; 2228 A; 31 C; 950 G; 2836 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 58.2;  DB 6;  Length 6045;
  Best Local Similarity   44.6%;  Pred. No. 0.084;
  Matches  283;  Conservative    0;  Mismatches  343;  Indels    9;  Gaps    1;

Qy        856 TATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGTGATATATAATCTAAC 915
Db       3012 TAAGATATATTAAAATTTTTGATATTTTTTTGAAATATTATTTATTAATTATTATTTATA 3071

Qy        916 TTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTGACCTCAGTCCCTAGT 975
Db       3072 TGAAAGGTTATTTAAAGTATTTTGAGTTATATAATTTTTTGTTGATTTATATAGTAATT 3131

Qy        976 TAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGCAA 1035
Db       3132 AAATTATTATATGTTAAGTAATAATAAATTTTAGAAATTTTTATTTTGGATTATTTTAAA 3191

Qy       1036 CTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCGTATTA 1095
Db       3192 TAAGGATTTTAGTGATTATTTTATATATTTTTTTTTTAGGTTAGTTATTTTTTAAAAATA 3251

Qy       1096 AATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAAT 1155
Db       3252 AAAATGTTAAT---------ATAAGGTTAGAGTTTTTATATGATAATAAGTAGATAATTT 3302

Qy       1156 GTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTA 1215
Db       3303 TATATATTAAATGTGTTTTGTAGATAAGTTATAGATACGATTTTTTTTTAAAGAAGTTT 3362

Qy       1216 TATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGT 1275
Db       3363 AAAAATTTAGATAGTTATATTTTTATTTTTTTGAGAAAAATATAGAGTATATTTTTTTATT 3422

Qy       1276 TGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAG 1335
Db       3423 ATTTGTTTTATAAATGGAGTTATTTATAAATTATGGAGTAATAATTAAATGTTAATGATA 3482

Qy       1336 AAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTTTTCGTAACCACAG 1395
Db       3483 AATATTTTTTTTTATTATATAGGAATTTTTTATGATTTTTAAATTTTTTATAGTATGTTTTA 3542

Qy       1396 TTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAAAAGATCGTTTTTGTA 1455
Db       3543 GTTTAGTTTGTTAAAGGTAATGTGAAGGTTTACGAATAATTTAGGAATCGAGTTTATATG 3602

Qy       1456 AAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGT 1490
Db       3603 GTTTTTTGAAGTTAAATTTATAAATTATTTTTTTT 3637

RESULT 14
ABZ10100
ID   ABZ10100 standard; DNA; 8056 BP.
XX
AC   ABZ10100;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 240; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related DNA
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 58.2;  DB 7;  Length 8056;
  Best Local Similarity   45.0%;  Pred. No. 0.088;
  Matches  259;  Conservative    0;  Mismatches  313;  Indels    3;  Gaps    1;

Qy        830 AAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTG 889
Db       6875 AAAAAAAATCGTTAATTTTTTTTTTTCGAATAAATCGAACGTTTAATTTTTTTTTAAAT 6934

Qy        890 TCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTA 949
Db       6935 TTTTTTTTTTTTTTATTATTAAAAATACGATTATATATAAAAAATATATCGATTTTAAT 6994

Qy        950 TGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAA 1009
Db       6995 TAAAAAAAAAAAAAAAAATATTATTTTCGTATTTTTTTTAAAAAAAAAAAAAATTTTAAA 7054

Qy       1010 AATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAA 1069
Db       7055 TTAATTAAAATTTAATTATTTTTTAAATATTAATTTTTTAATTTATTAATTTATAATTTT 7114

Qy       1070 TATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGA 1129
Db       7115 TTTTAAATATTAATAAAAATATATTATATATAATTCGTAAACGTATATTTATATTTTATT 7174

Qy       1130 TAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTG 1189
Db       7175 TACGTTTCGATTTAATTAATTATTTAATTTTATATATAATTTATTTATTTTTTAAAATAA 7234

Qy       1190 TATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTA---GAGTT 1246
```

```
Db       7235 TTTATATTTAATATATAATATATATTTTTTCGAAATTTTTTATTTTATATTATTTTACGA 7294

Qy       1247 TATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCG 1306

Db       7295 TATAAAAAATAAAAAATATTTTTTTTATTTTTAATTTTTTATTTAAAAATATTTTAAATA 7354

Qy       1307 TAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGT 1366

Db       7355 AATTTTTTTTTTTTTTTTATTATTTTTAAAAAAAAAAATAATTCGAAAAAAAAAAATACGT 7414

Qy       1367 GTAATTCATAGTCTTTTTCGTAACCACAGTTCTAT 1401

Db       7415 ATATAATATATACGTTATTTAAAATAAATTATTAT 7449
```

RESULT 15
ABQ66998
ID ABQ66998 standard; DNA; 37515 BP.
XX
AC ABQ66998;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 28.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 28; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 37515 BP; 11428 A; 425 C; 7165 G; 18497 T; 0 U; 0 Other;

Query Match 2.7%; Score 57.6; DB 6; Length 37515;
Best Local Similarity 48.8%; Pred. No. 0.15;
Matches 156; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

```
Qy       1146 AACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTG 1205

Db       6823 AAGTATAAGTTTTTAGAGATATTTTTAGTTTTTTTTTGTAGTAAAGTTAATAAAATTTTA 6882

Qy       1206 AAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTG 1265

Db       6883 ATAATATTTAAAGGATGTTTATGGTTATTTGGTAAAAAGAATAATTATAATTGATAAAGA 6942

Qy       1266 ATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCATCC 1325

Db       6943 TTGTAGTATTTGATATTTAGAATTAATTATTTTGGATTTTTTTGTTATTTTTAAATAAAA 7002

Qy       1326 TCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTTTTC 1385

Db       7003 TGATTTGTGAGAAATGTATTTTTTTTTAGGTAATATTTTTTTTTTAATTGTATTTATTAT 7062

Qy       1386 GTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAAAAGAT 1445

Db       7063 ATAAATTAAATTAAATTTGAATAATCGTATAAAGTTATTTTTTTAAAAAATAAAATAAGT 7122

Qy       1446 CGTTTTTGTAAAATTTAGTA 1465

Db       7123 AAATATTGTTGTAATTAATA 7142
```

Search completed: July 31, 2004, 04:28:03
Job time : 866 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 04:05:28 ; Search time 163 Seconds
(without alignments)
7316.500 Million cell updates/sec

Title: US-10-027-880-1
Perfect score: 2149
Sequence: 1 gtcgaattgtgatatattgt.....actaattgctaattaatcag 2149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	59.4	2.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
	2	59.4	2.8	8920	3	US-09-150-741-1	Sequence 1, Appli
c	3	51.4	2.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
	4	51.2	2.4	6124	4	US-08-213-419B-3	Sequence 3, Appli
	5	50.8	2.4	6317	4	US-10-204-708-11	Sequence 11, Appl
	6	50.2	2.3	19513	4	US-10-204-708-40	Sequence 40, Appl
	7	50	2.3	5455	4	US-10-204-708-33	Sequence 33, Appl
c	8	49.2	2.3	640681	4	US-09-790-988-1	Sequence 1, Appli
	9	48.6	2.3	11049	4	US-10-204-708-23	Sequence 23, Appl
	10	48.4	2.3	6765	4	US-09-620-312D-481	Sequence 481, App
	11	47.8	2.2	5562	4	US-10-204-708-63	Sequence 63, Appl
c	12	47.8	2.2	6243	2	US-09-056-075-1	Sequence 1, Appli
	13	47.4	2.2	8961	4	US-10-204-708-80	Sequence 80, Appl
	14	47.2	2.2	6156	4	US-10-204-708-60	Sequence 60, Appl
	15	46.8	2.2	6669	4	US-10-204-708-6	Sequence 6, Appli
	16	46.6	2.2	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	17	45.8	2.1	5340	4	US-09-627-122-21	Sequence 21, Appl
c	18	45.8	2.1	168575	4	US-09-426-290-1	Sequence 1, Appli
	19	45.8	2.1	640681	4	US-09-790-988-1	Sequence 1, Appli
	20	45.6	2.1	53332	4	US-09-801-861-3	Sequence 3, Appli
	21	44.6	2.1	4185	4	US-09-417-485D-7	Sequence 7, Appli
c	22	44.4	2.1	5852	1	US-07-867-106-2	Sequence 2, Appli
	23	44.2	2.1	731	1	US-08-451-405A-2	Sequence 2, Appli
	24	44.2	2.1	15016	4	US-09-601-198-60	Sequence 60, Appl
c	25	44	2.0	1887	4	US-09-601-198-4	Sequence 4, Appli
	26	43.8	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	27	43.8	2.0	10467	4	US-10-204-708-1	Sequence 1, Appli
	28	43.6	2.0	636	3	US-08-998-416-1137	Sequence 1137, Ap
	29	43.6	2.0	6113	4	US-10-204-708-14	Sequence 14, Appl
c	30	43.6	2.0	6317	4	US-10-204-708-12	Sequence 12, Appl
	31	43.6	2.0	11049	4	US-10-204-708-22	Sequence 22, Appl
	32	43.2	2.0	615	3	US-08-998-416-186	Sequence 186, App
c	33	43.2	2.0	6866	4	US-10-204-708-20	Sequence 20, Appl
c	34	43	2.0	860	3	US-08-998-416-287	Sequence 287, App
	35	43	2.0	5152	4	US-10-204-708-74	Sequence 74, Appl
c	36	43	2.0	392000	4	US-10-027-983-11	Sequence 11, Appl
	37	42.8	2.0	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	38	42.8	2.0	2749	1	US-08-344-695-1	Sequence 1, Appli
c	39	42.8	2.0	5152	4	US-10-204-708-47	Sequence 47, Appl
c	40	42.8	2.0	6124	4	US-08-213-419B-3	Sequence 3, Appli
	41	42.8	2.0	7304	4	US-10-204-708-44	Sequence 44, Appl
c	42	42.6	2.0	832	4	US-09-621-976-2813	Sequence 2813, Ap
	43	42.6	2.0	6306	4	US-10-204-708-50	Sequence 50, Appl
c	44	42.6	2.0	786431	4	US-09-751-389-3	Sequence 3, Appli
	45	42.4	2.0	5152	4	US-10-204-708-73	Sequence 73, Appl

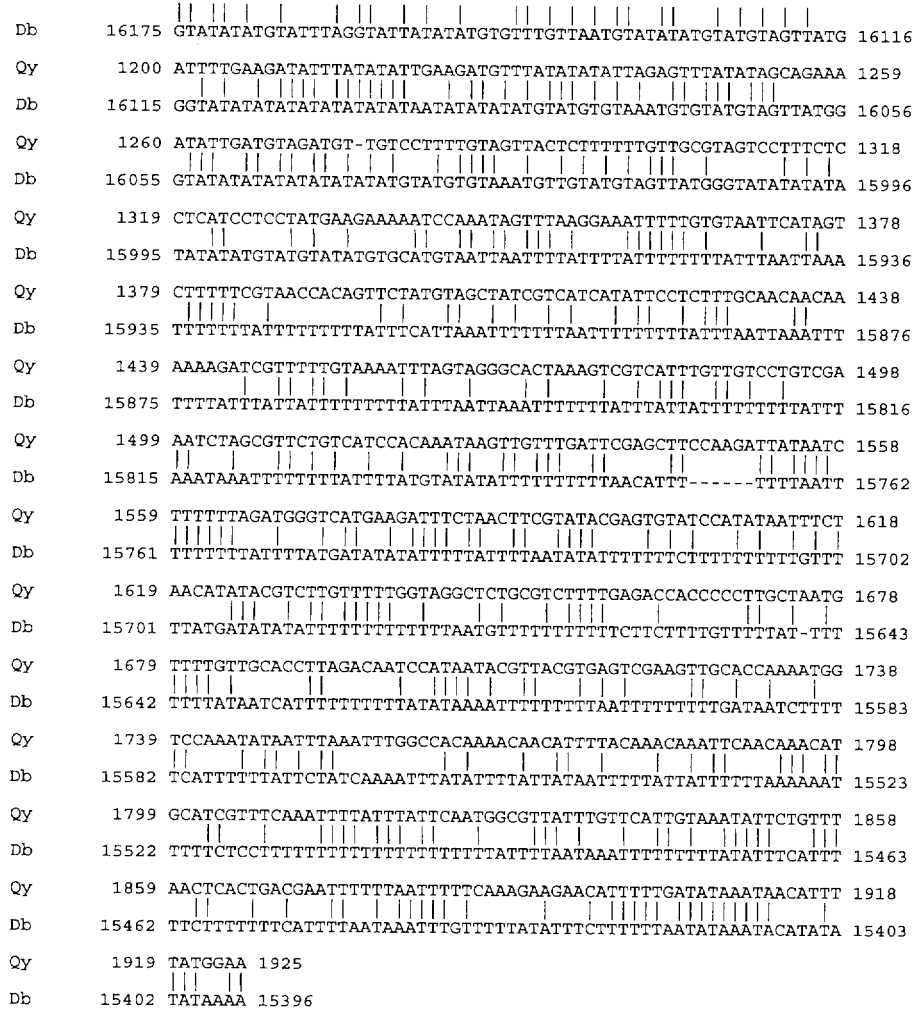
ALIGNMENTS

RESULT 1
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29.009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 2.8%; Score 59.4; DB 2; Length 8920;
Best Local Similarity 45.7%; Pred. No. 0.00056;
Matches 207; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

Qy 266 ATATATACTGATATTGATATGCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATC 325
Db 333 ATATATATATATATGTTTGTGTGTTCATTTGTTTATAAAATTACTTGAAATATAAAACTT 392

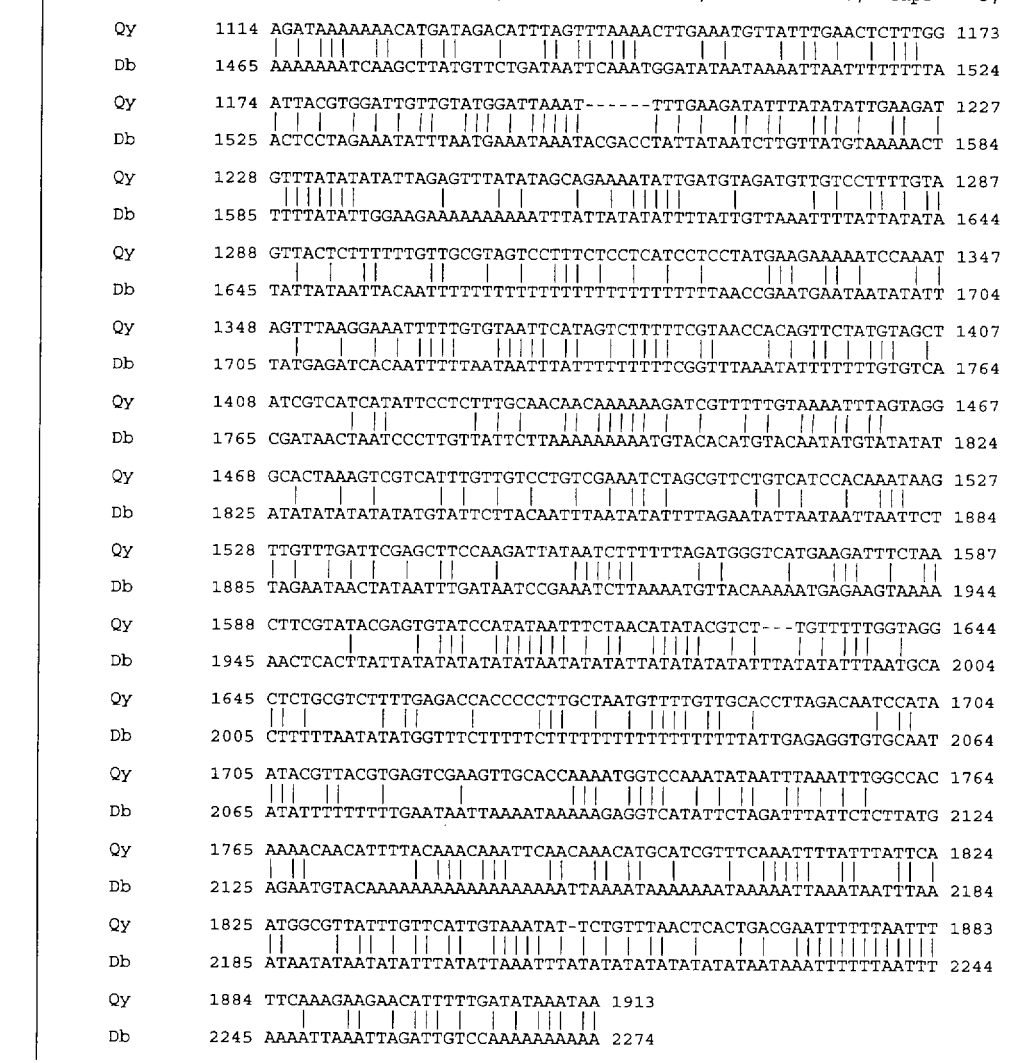
```
Qy        326 AGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAA 385
Db        393 ATTAATATATTTCCAATTAATATGAATACAATTATTAATATTTTGATGTGTACACATTAA 452

Qy        386 AAAAGTGGTATACTGGTATATACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGG 445
Db        453 TATAGTTTTACACTTCTTATAATAAAACCATCCTATATATTATACACAATATATAATACT 512

Qy        446 AGAAAATGGACGTGTCTCTTTGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAA 505
Db        513 CCCCAATATTGTGGTTCCTATAATTTTATTTATATATTTATTTATTAATTTATTCATTTA 572

Qy        506 ACACTTCGATGTCTCGCTTAAATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTA 565
Db        573 TTTATTTTTTTTCTTAGTTTATAAAATAGTAATTCTACTAATTTAAAAAAAAAAAAAAAA 632

Qy        566 AAGAAAATTCGTCGAAGCCACAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCAT 625
Db        633 AAAAAAAAAAAAAGAAAAAAAAAAAATTTACATATGAAAAATGAACTTGTATATGTAAAT 692

Qy        626 CAAAAATATATTAATAGAAAAGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAA 685
Db        693 TTATAAATATTTTAAACATAAATATAAATGTATAAAAAAAAAAAAGAAAAATGGGAAAAA 752

Qy        686 AAAAGATAAGAAGGTAGGCAGAAGAAAACGTAT 718
Db        753 ATAATATAGATATATATATAAATATATATATAT 785


RESULT 2
US-09-150-741-1
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
;  APPLICANT: Stewart et al.
;  TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
;  TITLE OF INVENTION: Synthetase II
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/150,741
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PL6380
;  EARLIER FILING DATE: 1992-12-16
;  EARLIER APPLICATION NUMBER: AU93/00617
;  EARLIER FILING DATE: 1993-12-02
;  EARLIER APPLICATION NUMBER: 08/446,855
;  EARLIER FILING DATE: 1995-07-06
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 8920
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-150-741-1

  Query Match             2.8%;  Score 59.4;  DB 3;  Length 8920;
  Best Local Similarity   45.7%;  Pred. No. 0.00056;
  Matches  207;  Conservative    0;  Mismatches  246;  Indels    0;  Gaps    0;

Qy        266 ATATATACTGATATTGATATGCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATC 325
Db        333 ATATATATATATATGTTTGTGTGTTCATTTGTTTATAAAATTACTTGAAATATAAAACTT 392

Qy        326 AGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAA 385
Db        393 ATTAATATATTTCCAATTAATATGAATACAATTATTAATATTTTGATGTGTACACATTAA 452

Qy        386 AAAAGTGGTATACTGGTATATACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGG 445
Db        453 TATAGTTTTACACTTCTTATAATAAAACCATCCTATATATTATACACAATATATAATACT 512

Qy        446 AGAAAATGGACGTGTCTCTTTGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAA 505
Db        513 CCCCAATATTGTGGTTCCTATAATTTTATTTATATATTTATTTATTAATTTATTCATTTA 572

Qy        506 ACACTTCGATGTCTCGCTTAAATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTA 565
Db        573 TTTATTTTTTTTCTTAGTTTATAAAATAGTAATTCTACTAATTTAAAAAAAAAAAAAAAA 632

Qy        566 AAGAAAATTCGTCGAAGCCACAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCAT 625
Db        633 AAAAAAAAAAAAAGAAAAAAAAAAAATTTACATATGAAAAATGAACTTGTATATGTAAAT 692

Qy        626 CAAAAATATATTAATAGAAAAGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAA 685
Db        693 TTATAAATATTTTAAACATAAATATAAATGTATAAAAAAAAAAAAGAAAAATGGGAAAAA 752

Qy        686 AAAAGATAAGAAGGTAGGCAGAAGAAAACGTAT 718
Db        753 ATAATATAGATATATATATAAATATATATATAT 785


RESULT 3
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13

  Query Match             2.4%;  Score 51.4;  DB 2;  Length 19124;
  Best Local Similarity   44.2%;  Pred. No. 0.041;
  Matches  348;  Conservative    0;  Mismatches  431;  Indels    8;  Gaps    3;

Qy       1140 GTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAA 1199
```



```
Db      16175 GTATATATGTATTTAGGTATTATATATGTGTTTGTTAATGTATATATGTATGTAGTTATG 16116

Qy       1200 ATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAA 1259

Db      16115 GGTATATATATATATATATATAATATATATATGTATGTGTAAATGTGTATGTAGTTATGG 16056

Qy       1260 ATATTGATGTAGATGT-TGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTC 1318

Db      16055 GTATATATATATATATATATGTATGTGTAAATGTTGTATGTAGTTATGGGTATATATATA 15996

Qy       1319 CTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGT 1378

Db      15995 TATATATGTATGTATATGTGCATGTAATTAATTTTATTTTATTTTTTTTATTTAATTAAA 15936

Qy       1379 CTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAA 1438

Db      15935 TTTTTTTATTTTTTTTTATTTCATTAAATTTTTTTAATTTTTTTTTATTTAATTAAATTT 15876

Qy       1439 AAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGA 1498

Db      15875 TTTTATTTATTATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTATTT 15816

Qy       1499 AATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATC 1558

Db      15815 AAATAAATTTTTTTTATTTTATGTATATATTTTTTTTTTTAACATTT------TTTTAATT 15762

Qy       1559 TTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCT 1618

Db      15761 TTTTTTTATTTTATGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTTGTTT 15702

Qy       1619 AACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATG 1678

Db      15701 TTATGATATATATTTTTTTTTTTTTAATGTTTTTTTTTTTCTTCTTTTGTTTTTAT-TTT 15643

Qy       1679 TTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGG 1738

Db      15642 TTTTATAATCATTTTTTTTTTATATAAAATTTTTTTTTAATTTTTTTTTGATAATCTTTT 15583

Qy       1739 TCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACAT 1798

Db      15582 TCATTTTTTATTCTATCAAAATTTATATTTTATTATAATTTTTATTATTTTTTAAAAAAT 15523

Qy       1799 GCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTT 1858

Db      15522 TTTTCTCCTTTTTTTTTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTATATTTCATTT 15463

Qy       1859 AACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTT 1918

Db      15462 TTCTTTTTTTCATTTTAATAAATTTGTTTTTATATTTCTTTTTTAATATAAATACATATA 15403

Qy       1919 TATGGAA 1925

Db      15402 TATAAAA 15396
```

RESULT 4
US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: JII-002CNCP
; CURRENT APPLICATION NUMBER: US/08/213,419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match 2.4%; Score 51.2; DB 4; Length 6124;
Best Local Similarity 44.1%; Pred. No. 0.033;
Matches 357; Conservative 0; Mismatches 443; Indels 10; Gaps 3;



```
Qy       1114 AGATAAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGG 1173

Db       1465 AAAAAAATCAAGCTTATGTTCTGATAATTCAAATGGATATAATAAAATTAATTTTTTTTA 1524

Qy       1174 ATTACGTGGATTGTTGTATGGATTAAAT------TTTGAAGATATTTATATATTGAAGAT 1227

Db       1525 ACTCCTAGAAATATTTAATGAAATAAATACGACCTATTATAATCTTGTTATGTAAAAACT 1584

Qy       1228 GTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTA 1287

Db       1585 TTTTATATTGGAAGAAAAAAAAAATTTATTATATATTTTATTGTTAAATTTTATTATATA 1644

Qy       1288 GTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAAT 1347

Db       1645 TATTATAATTACAATTTTTTTTTTTTTTTTTTTTTTTTTTTTAACCGAATGAATAATATATT 1704

Qy       1348 AGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCT 1407

Db       1705 TATGAGATCACAATTTTTAATAATTTATTTTTTTTTCGGTTTAAATATTTTTTTGTGTCA 1764

Qy       1408 ATCGTCATCATATTCCTCTTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGG 1467

Db       1765 CGATAACTAATCCCTTGTTATTCTTAAAAAAAAATGTACACATGTACAATATGTATATAT 1824

Qy       1468 GCACTAAAGTCGTCATTTGTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAG 1527

Db       1825 ATATATATATATATGTATTCTTACAATTTAATATATTTTAGAATATTAATAATTAATTCT 1884

Qy       1528 TTGTTTGATTCGAGCTTCCAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAA 1587

Db       1885 TAGAATAACTATAATTTGATAATCCGAAATCTTAAAATGTTACAAAAATGAGAAGTAAAA 1944

Qy       1588 CTTCGTATACGAGTGTATCCATATAATTTCTAACATATACGTCT---TGTTTTTGGTAGG 1644

Db       1945 AACTCACTTATTATATATATATAATATATTATATATATATTTATATATTTAATGCA 2004

Qy       1645 CTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATA 1704

Db       2005 CTTTTTAATATATGGTTTCTTTTTCTTTTTTTTTTTTTTTTTTTATTGAGAGGTGTGCAAT 2064

Qy       1705 ATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCAC 1764

Db       2065 ATATTTTTTTTTGAATAATTAAAATAAAAAGAGGTCATATTCTAGATTTATTCTCTTATG 2124

Qy       1765 AAAACAACATTTTACAAACAAATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCA 1824

Db       2125 AGAATGTACAAAAAAAAAAAAAAAAAAATTAAAATAAAAAAATAAAAATTAAATAATTTAA 2184

Qy       1825 ATGGCGTTATTTGTTCATTGTAAATAT-TCTGTTTAACTCACTGACGAATTTTTTAATTT 1883

Db       2185 ATAATATAATATATTTATATTAAATTTATATATATATATATATAATAAATTTTTTAATTT 2244

Qy       1884 TTCAAAGAAGAACATTTTTGATATAAATAA 1913

Db       2245 AAAATTAAATTAGATTGTCCAAAAAAAAAA 2274
```

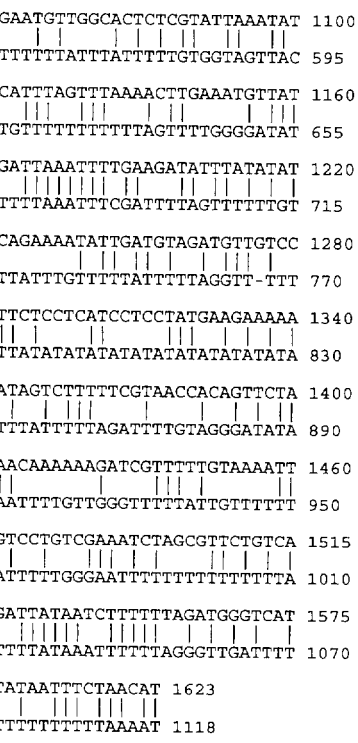
```
RESULT 5
US-10-204-708-11
; Sequence 11, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 11
;   LENGTH: 6317
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-11

  Query Match             2.4%;  Score 50.8;  DB 4;  Length 6317;
  Best Local Similarity   43.5%;  Pred. No. 0.041;
  Matches  229;  Conservative    0;  Mismatches  297;  Indels    0;  Gaps    0;

Qy        776 TGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCA 835

Db       2549 TGATTTTTTTTTATATAGTTATTGATGTTTAGTTTTTTTTATTTAGTATAGTGTTTTTAA 2608

Qy        836 TATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTA 895

Db       2609 GATTTATTTATATTGTATATTAGTAGTTTGTTTTTATTGTTGAAGAGTATTGTATTGTAT 2668

Qy        896 TTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTC 955

Db       2669 TGAGGTATATTTGTTTATTTATTTATTAGGATAGATGTATATGTTAAGTTGTGGGGTGGG 2728

Qy        956 TACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTG 1015

Db       2729 AATGAATATAGATTAATTTTTAAGTGTTAGTTTAGTTGTGAATATTTTTGTGTTAATATT 2788

Qy       1016 AAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGA 1075

Db       2789 TAATGGAAGATATTTTGGTAGTAAATATTTATATTATAGTTATTTGATGGAAATTAAAGA 2848

Qy       1076 ATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACA 1135

Db       2849 TTTAAAATAAAATATTTTTTGAAATAGTTTTAAGATTTATAAATTTTAAATTATGTATTT 2908

Qy       1136 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1195

Db       2909 AAAATTATAAGTAACGAAAAGTTTTTTTATTTTTTTTTTAAGTTAAATTTTTGGATATT 2968

Qy       1196 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1255

Db       2969 TTATGTTTTTATATATATATTTAATTATGTATATGTTAATATTTTTATATATTGATATGT 3028

Qy       1256 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTG 1301

Db       3029 AAATATATATATGTATATATATTTTTTAATTAGGTAAGTAGTTGTG 3074
```

```
RESULT 6
US-10-204-708-40
; Sequence 40, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 40
;   LENGTH: 19513
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-40

  Query Match             2.3%;  Score 50.2;  DB 4;  Length 19513;
  Best Local Similarity   54.6%;  Pred. No. 0.076;
  Matches  100;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       1120 AAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACG 1179

Db      10534 ATATAATTTATTTATTTAAAGTATATATAGTTTAATGGTATTTAGTATGTTTAGAGTTAG 10593

Qy       1180 TGGATTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATAT 1239

Db      10594 GTAATTATTATTATAGTTAATTTTAGAATTTTTTTATTATTTAAAAAGTTTAAATTTTT 10653

Qy       1240 TAGAGTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTT 1299

Db      10654 TAGTTTTTTTTTTGTATTTAGTTTATTTGTTTTTATTTTTGGTAGTTATTTATTTAATTT 10713

Qy       1300 TGT 1302

Db      10714 TTT 10716
```

```
RESULT 7
US-10-204-708-33
; Sequence 33, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
```

; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 33
; LENGTH: 5455
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-33

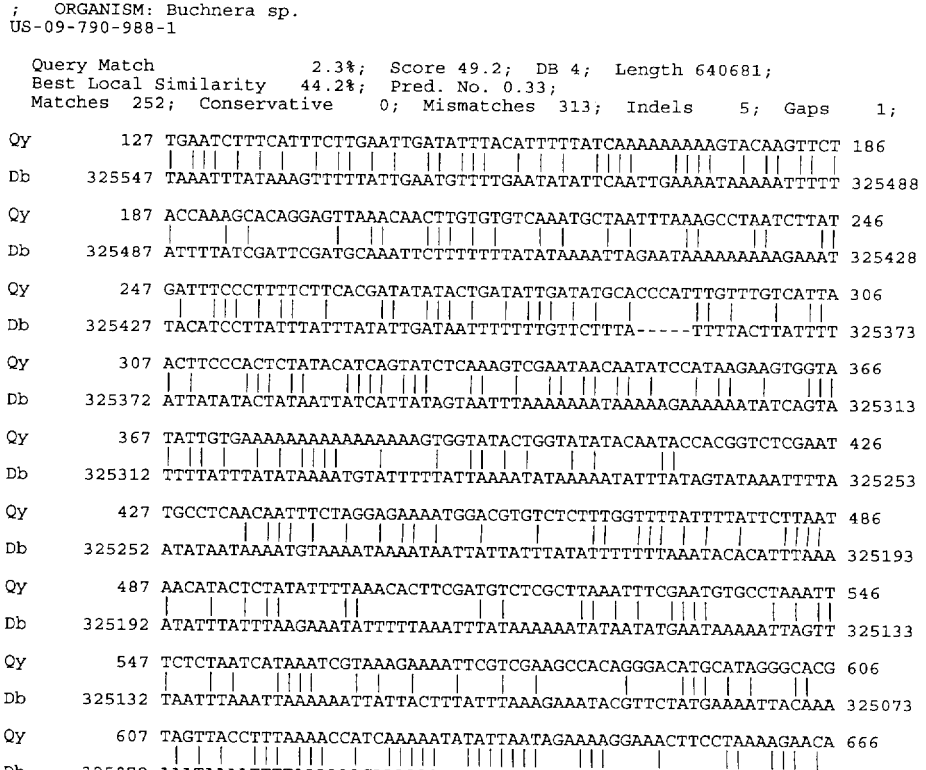
Query Match 2.3%; Score 50; DB 4; Length 5455;
Best Local Similarity 46.4%; Pred. No. 0.059;
Matches 273; Conservative 0; Mismatches 305; Indels 10; Gaps 3;



RESULT 8
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07

; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 2.3%; Score 49.2; DB 4; Length 640681;
Best Local Similarity 44.2%; Pred. No. 0.33;
Matches 252; Conservative 0; Mismatches 313; Indels 5; Gaps 1;



RESULT 9
US-10-204-708-23
; Sequence 23, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
; LENGTH: 11049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-23

Query Match 2.3%; Score 48.6; DB 4; Length 11049;
Best Local Similarity 44.2%; Pred. No. 0.15;
Matches 247; Conservative 0; Mismatches 309; Indels 3; Gaps 1;

Qy 1086 TCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAA 1145
Db 8407 TTTTTTTTTTAAAAAGAAATTTTTATGTGTGTATAAATAAGAAAAATGAAAATAGAATTG 8466

Qy 1146 AACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTG 1205
Db 8467 TTAATTGTTAGTTAGTATAGATATTTTGAGTTGGAATTTTGTTTTTTTTGTTGGAAAGTG 8526

Qy 1206 AAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTG 1265
Db 8527 AGGTGAGTTGAGTTTTTTTTTTGAAAATAATTAGTAGGTTTTAAAAGAAGGAATTTTTTT 8586

Qy 1266 ATGTAGATGTTGTCCTTT---TGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCA 1322
Db 8587 ATTGTGTTGTTTATTTTTTAATTTTTTTAATGTTAGTTTTTTCGTATTTTTTAAAATGAT 8646

Qy 1323 TCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTT 1382
Db 8647 TTTAGGTATTTGTTAGAGTTTTGGGAATTTTGGAATTGTTTATTTTGTTGATGTGTTTAT 8706

Qy 1383 TTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAAAA 1442
Db 8707 ATGTTATTTAGATTTTTTTTTGAGTTTAGATTTGTATTTTTGATTGTATGTTGTAAATGG 8766

Qy 1443 GATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAATC 1502
Db 8767 TATTGATTATATGTAAGTTGTTTGTTAGTTTTGTTTTTTATTTGTTTTTTTTTTGAGAGT 8826

Qy 1503 TAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTTTT 1562
Db 8827 TTTATTTTTGTGATTGAATTTATTGGTGATTTATTTGTTTAGATTAAATTTTTTATTATT 8886

Qy 1563 TTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAACA 1622
Db 8887 TTAGATTTAGTTTTGAATTTTATTTTTTTTTAGTAGATATTGAGTATTTTTTATATATTA 8946

Qy 1623 TATACGTCTTGTTTTTGGT 1641
Db 8947 GATTATGTATGGTTTATGT 8965

RESULT 10
US-09-620-312D-481
; Sequence 481, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 481
; LENGTH: 6765
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(3105)
US-09-620-312D-481

Query Match 2.3%; Score 48.4; DB 4; Length 6765;
Best Local Similarity 49.2%; Pred. No. 0.14;
Matches 127; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 994 TGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTC 1053
Db 5350 TTAATCTTACCTAAAAGTACTGAAAAATGCTTTTTAGTACTTTTTCATATTTTATACATT 5409

Qy 1054 ATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCT 1113
Db 5410 TATTTTCCGAATGTATCATTGAATAATTTTATTGAGTTATAAAAGTATCTTATTGCTATT 5469

Qy 1114 AGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGG 1173
Db 5470 TAATAAAAAATTAACACATAAAATGACTTGAATTGTCATCATTCTTTTTAAGATATTTAG 5529

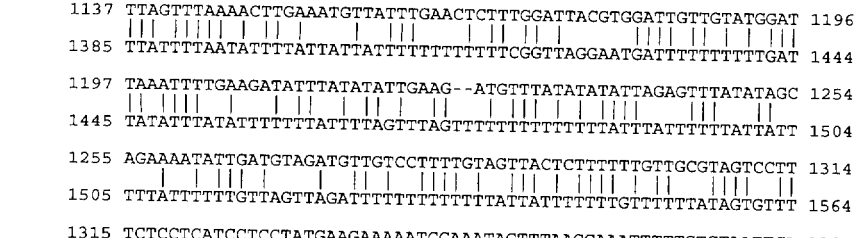
Qy 1174 ATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTAT 1233
Db 5530 TTAAACTGACTTAATGTATGGCCTTCAATTTTTTGTGTCCTTATTTTTCTGATCATTTC 5589

Qy 1234 ATATATTAGAGTTTATAT 1251
Db 5590 TCCTTTTATAGTTTACAT 5607

RESULT 11
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1

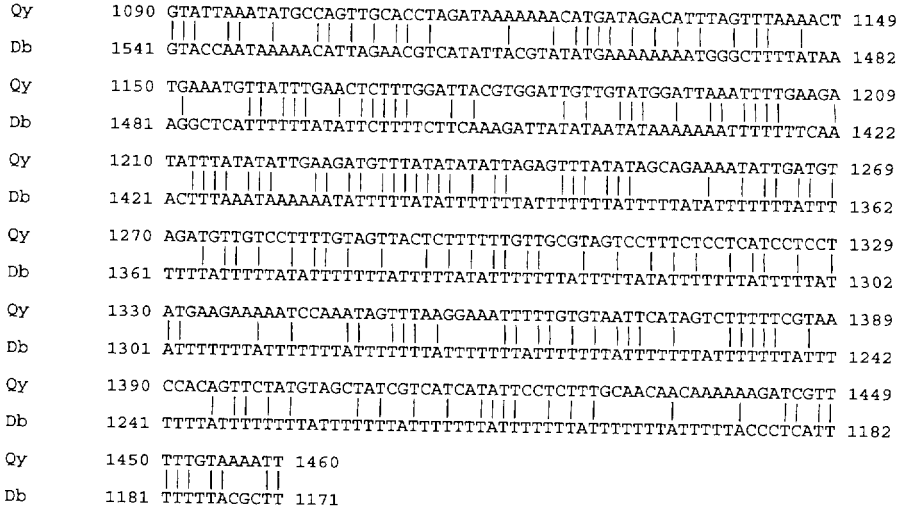
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
; LENGTH: 5562
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

Query Match 2.2%; Score 47.8; DB 4; Length 5562;
Best Local Similarity 48.4%; Pred. No. 0.18;
Matches 195; Conservative 0; Mismatches 202; Indels 6; Gaps 2;



RESULT 12
US-09-056-075-1/c
; Sequence 1, Application US/09056075
; Patent No. 5955368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,075
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3770..4013
; OTHER INFORMATION: /note= "RP4 origin of DNA transfer (oriT) from
; OTHER INFORMATION: plasmid RP4"
US-09-056-075-1

Query Match 2.2%; Score 47.8; DB 2; Length 6243;
Best Local Similarity 45.6%; Pred. No. 0.19;
Matches 169; Conservative 0; Mismatches 202; Indels 0; Gaps 0;



RESULT 13
US-10-204-708-80
; Sequence 80, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 80
; LENGTH: 8961
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (5096)
; OTHER INFORMATION: n is a or g or c or t
US-10-204-708-80

Query Match 2.2%; Score 47.4; DB 4; Length 8961;
Best Local Similarity 46.3%; Pred. No. 0.26;
Matches 156; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

```
Qy       972 TAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCT 1031
Db      5793 TCGTTTTATAGATGGATATTTAATTTTTTTTTTTTAATATTGGAAAGTTATTTTATAAT 5852

Qy      1032 GCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCGT 1091
Db      5853 AATTTTATTTTTTTATTTATATTTTGATATTTTATAATAATTGTAAGGAGTTCGTGGTTA 5912

Qy      1092 ATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTG 1151
Db      5913 ATTAAATATATTATAATATTTTTAATATAATTTAGTGTTATATTTTATAATGTTTTATTT 5972

Qy      1152 AAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATA 1211
Db      5973 TTATTTTATTATTTTTATATTTTTCGAGTTAGGTATTTTATAATATTTATATTTTTTTTA 6032

Qy      1212 TTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAG 1271
Db      6033 TTTTTATTTTTAGGCGTTGTTAGTTTATTTTAATATTTAGAGTAGGGAGGGTTAGGTAGG 6092

Qy      1272 ATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTA 1308
Db      6093 TTATTTACGTTTTTTTGTTTCGTTTTTATTTGTATA 6129
```

RESULT 14
US-10-204-708-60
; Sequence 60, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 60
; LENGTH: 6156
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-204-708-60

Query Match 2.2%; Score 47.2; DB 4; Length 6156;
Best Local Similarity 44.1%; Pred. No. 0.26;
Matches 240; Conservative 0; Mismatches 303; Indels 1; Gaps 1;

```
Qy       765 GAGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAG 824
Db      4302 GAGGGGGTGGGGATTTTTTTTTTTTAAAAATAATTTAGAATATAAATTTTTAGTTTGAGA 4361

Qy       825 GGTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTG 884
Db      4362 AAAATATTATTTGAAGATTGGTTATAAATATATTTGAAAGGAAAATTTAGGTATTTTTTA 4421

Qy       885 CTTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCT 944
Db      4422 AGGTGTTTTTTTTGATTTTATATAAGATTTATTTATTTATTTGTATTAAAAATTATTTTT 4481

Qy       945 ACCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACG 1004
Db      4482 TAGGAAGTGTTATAGTAATTTTTTAATATTTAAGTAATTTTTTAAATAAGTGGATATAGT 4541

Qy      1005 CCCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATG 1064
Db      4542 GTTATATGTGTATTAATTTATTTTGTATTAATTGATTGAAATTATATTTATTTAGTTATA 4601

Qy      1065 TTAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAA 1124
Db      4602 TTTTTTTTTGAATGAAATTATTGTTTTTTATTTTGGTTTTATTTATTTTTTATATAATAA 4661

Qy      1125 CATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTT-TGGATTACGTGGA 1183
Db      4662 GTTTATTTAATTATATTTAAGAGTTAGAATATTGATTTTAATTTTTAAATATTATTTTTA 4721

Qy      1184 TTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGA 1243
Db      4722 TTGATGAGAGTTGGTTGTATGGATTTAAATTAGAATGTAGTTTTTTTTTTTATATGTTAAA 4781

Qy      1244 GTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTT 1303
Db      4782 AGTAATTTAAAATTATATTAGCGATTTTTTTTTTTTAGAGTTTTGTTTTTTAATATTTTT 4841

Qy      1304 GCGT 1307
Db      4842 ACGT 4845
```

RESULT 15
US-10-204-708-6
; Sequence 6, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 6
; LENGTH: 6669

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-6

Query Match 2.2%; Score 46.8; DB 4; Length 6669;
Best Local Similarity 43.9%; Pred. No. 0.32;
Matches 339; Conservative 0; Mismatches 427; Indels 6; Gaps 3;

```
Qy       1117 TAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATT 1176
Db       3438 TTAAATATGGTTTTTTAGTTTTAGTATTGTTAATATTTTGGGTTTTATAATTTTTTATTG 3497

Qy       1177 ACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATA 1236
Db       3498 TGGGGAGTTATTTTGTGTTTAAATATTGTAGGATGTTTAGTAGTATTTTTGGTTTTTATT 3557

Qy       1237 TATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTT 1296
Db       3558 TATTAGA--TTATAGTAGGATATTCGTTTTTTATTTTATTTTTATTTAGTTGTGATTAGT 3615

Qy       1297 TTTTGTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGG 1356
Db       3616 GAAAATGTTTTTAGATATTGTTAGTATGTTTTTTGGAGGGTAGAATTATTTTTAGTTAAG 3675

Qy       1357 AAATTTTTGTGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATC 1416
Db       3676 AAAGTTTTAAGAAAGTAGTATTATTTTTTAATTGAAATAATTTATATCGATGTGTGAGTA 3735

Qy       1417 ATATTCCTCTTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAG 1476
Db       3736 TTTTTTTTTATAAATTAGTTATTTGTTGATTTTTTGTTTTTTATTTAATGGATAAAAA 3795

Qy       1477 TCGTCATTTGTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGAT 1536
Db       3796 ACGTTAGTTATGGTTTTTTTATTTTAAAAATTTTTGTTATTTATATTTTATTTTTATTGT 3855

Qy       1537 TCGAGCTTCCAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATA 1596
Db       3856 TATTAATTTCGTAATGTGTTTGTTATTATTTTTTTTTTGGATTTTTGTTATGGTTTTTTA 3915

Qy       1597 CGAGTGTATCCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTT 1656
Db       3916 GATAGTTGTTTTCGTAATTTATTTTATTTA-TTGTTGTTATAAATTTTTTTGAAGTTTTT 3974

Qy       1657 TGAGACCACCCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTG 1716
Db       3975 ATATTATATTTTTTAGTTTGAAAGTTGAGATAATTTTATTTTTATTAAAAATTTTAGATG 4034

Qy       1717 AGTCGAAGT---TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACA 1773
Db       4035 GGTAGGAGTATATAATATAAAAAAGTTTTTTTATGTTTAATGTTTTTATTTTTGTCGTTT 4094

Qy       1774 TTTTACAAACAAATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTA 1833
Db       4095 ATTATTTTATATTTTAAATAGATATATATCGTGTATATTTTTTTTTATATGAGGTTTATTA 4154

Qy       1834 TTTGTTCATTGTAAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTT 1885
Db       4155 TTTATTTATGATTTTTTTTTTATTAGTATTTTTTCGTTATTTTTAGTTTTTT 4206
```

Search completed: July 31, 2004, 08:23:31
Job time : 171 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 06:47:54 ; Search time 975 Seconds
(without alignments)
10807.024 Million cell updates/sec

Title: US-10-027-880-1
Perfect score: 2149
Sequence: 1 gtcgaattgtgatatattgt.....actaattgctaattaatcag 2149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

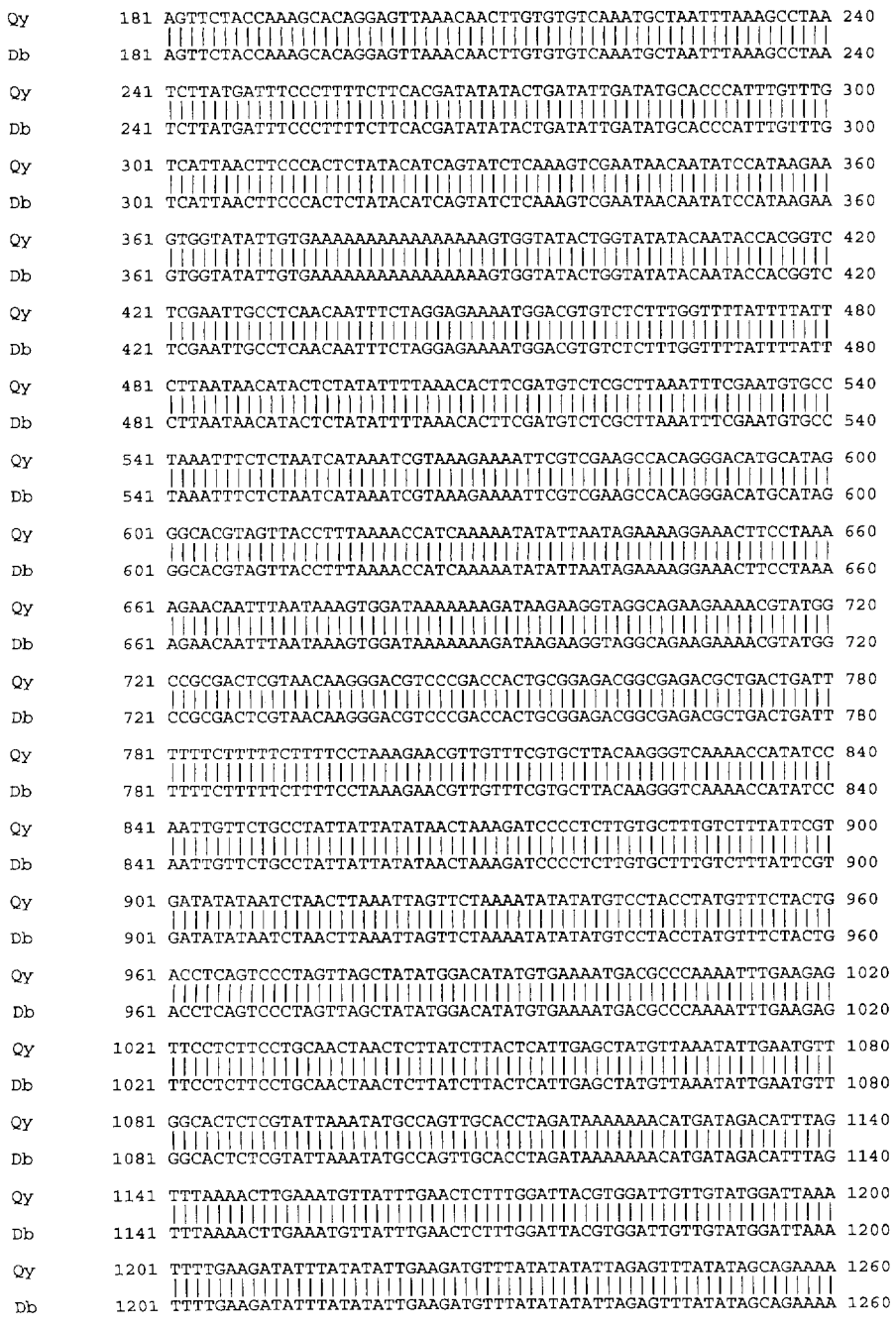
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2149	100.0	2149	15	US-10-027-880-1	Sequence 1, Appli
	2	2149	100.0	4280	15	US-10-027-880-2	Sequence 2, Appli
	3	2149	100.0	4309	15	US-10-027-880-4	Sequence 4, Appli
	4	2149	100.0	4413	15	US-10-027-880-3	Sequence 3, Appli
c	5	63.6	3.0	7008	13	US-10-221-714A-254	Sequence 254, App
c	6	60.6	2.8	396	17	US-10-021-323-17801	Sequence 17801, A
	7	60.2	2.8	6175	15	US-10-311-455-1280	Sequence 1280, Ap
	8	59.8	2.8	7597	15	US-10-311-455-986	Sequence 986, App
	9	59.6	2.8	5822	15	US-10-311-455-1070	Sequence 1070, Ap
	10	59.4	2.8	33053	17	US-10-433-793-35	Sequence 35, Appl
	11	58.6	2.7	5413	13	US-10-221-714A-418	Sequence 418, App
	12	57.6	2.7	37515	17	US-10-433-793-28	Sequence 28, Appl
	13	56.8	2.6	8277	15	US-10-311-455-1301	Sequence 1301, Ap
	14	56.6	2.6	6292	13	US-10-221-714A-461	Sequence 461, App
	15	56.6	2.6	3673778	15	US-10-312-841-1	Sequence 1, Appli
	16	56.6	2.6	3673778	15	US-10-312-841-2	Sequence 2, Appli
	17	56.4	2.6	15951	15	US-10-311-455-1653	Sequence 1653, Ap
	18	56.4	2.6	15951	15	US-10-240-485-133	Sequence 133, App
c	19	56.2	2.6	17280	13	US-10-221-714A-498	Sequence 498, App
c	20	56	2.6	495	17	US-10-021-323-2679	Sequence 2679, Ap
	21	56	2.6	37973	15	US-10-311-455-2170	Sequence 2170, Ap
	22	55.8	2.6	6076	13	US-10-221-714A-386	Sequence 386, App
c	23	55.8	2.6	3673778	15	US-10-312-841-2	Sequence 2, Appli
c	24	55.6	2.6	393	9	US-09-960-352-4582	Sequence 4582, Ap
	25	55.6	2.6	7072	13	US-10-221-613-348	Sequence 348, App
	26	55.4	2.6	12237	15	US-10-311-455-2331	Sequence 2331, Ap
	27	55.4	2.6	15224	13	US-10-221-613-223	Sequence 223, App
	28	55	2.6	11745	15	US-10-240-453-206	Sequence 206, App
c	29	55	2.6	113515	15	US-10-311-455-2147	Sequence 2147, Ap
	30	54.8	2.6	5297	13	US-10-221-714A-235	Sequence 235, App
	31	54.8	2.6	14919	13	US-10-221-714A-228	Sequence 228, App
	32	54.6	2.5	8244	13	US-10-221-714A-117	Sequence 117, App
	33	54.6	2.5	8244	15	US-10-311-455-965	Sequence 965, App
	34	54.6	2.5	8244	17	US-10-433-793-61	Sequence 61, Appl
	35	54.6	2.5	13131	17	US-10-240-589C-57	Sequence 57, Appl
	36	54.4	2.5	5273	15	US-10-311-455-848	Sequence 848, App
	37	54.2	2.5	4661	17	US-10-433-793-51	Sequence 51, Appl
	38	54.2	2.5	6191	15	US-10-311-455-1190	Sequence 1190, Ap
	39	54	2.5	6223	13	US-10-221-613-135	Sequence 135, App
	40	54	2.5	37515	17	US-10-433-793-27	Sequence 27, Appl
	41	53.8	2.5	7498	15	US-10-311-455-230	Sequence 230, App
	42	53.8	2.5	7851	15	US-10-311-455-1733	Sequence 1733, Ap
	43	53.6	2.5	6665	15	US-10-239-676-4	Sequence 4, Appli
	44	53.6	2.5	6665	15	US-10-311-455-56	Sequence 56, Appl
	45	53.6	2.5	6665	15	US-10-240-453-4	Sequence 4, Appli

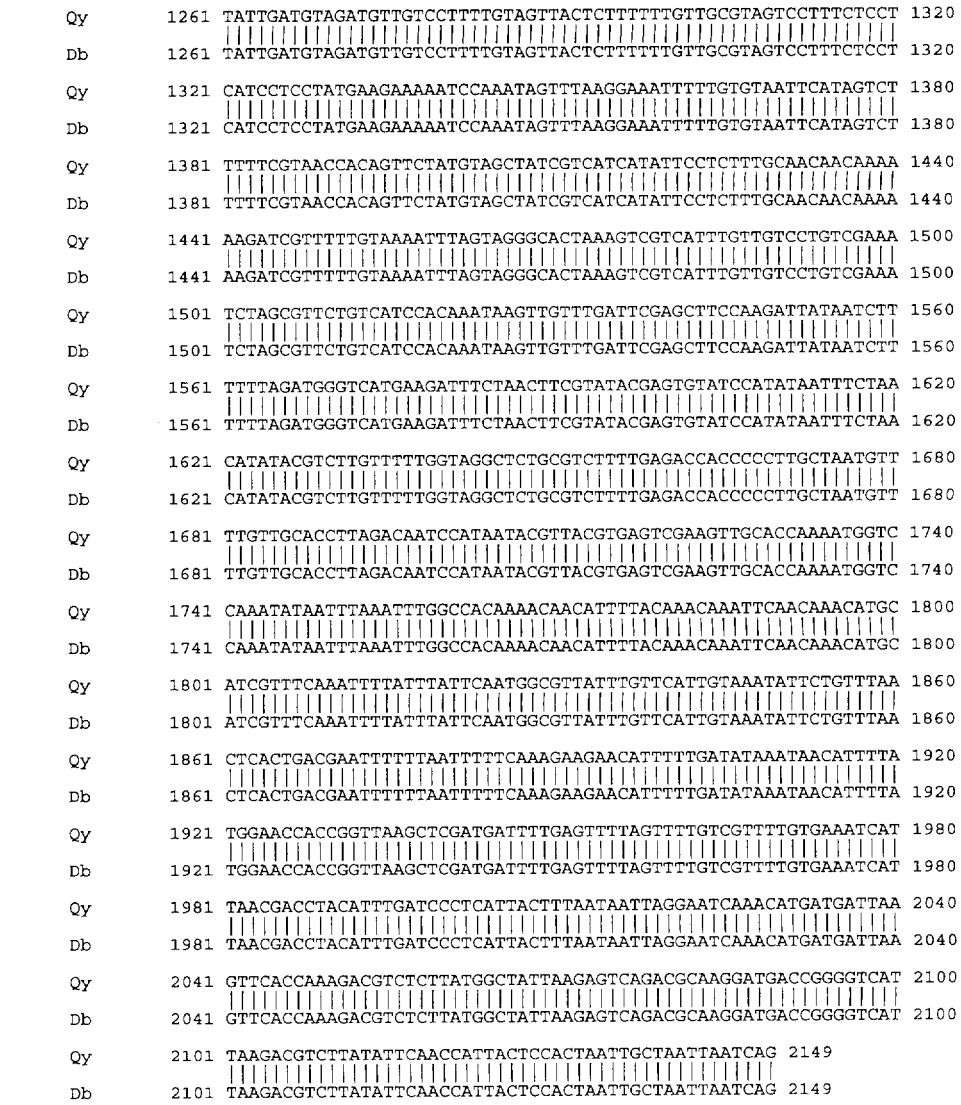
ALIGNMENTS

RESULT 1
US-10-027-880-1
; Sequence 1, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
; FILE REFERENCE: CHEP:003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2149
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-027-880-1

Query Match 100.0%; Score 2149; DB 15; Length 2149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
```





RESULT 2
US-10-027-880-2
; Sequence 2, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
; FILE REFERENCE: CHEP:003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4280
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Construction
; OTHER INFORMATION: promoter + sequence coding for the gus gene
US-10-027-880-2

Query Match 100.0%; Score 2149; DB 15; Length 4280;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy       1081 GGCACTCTCGTATTAAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCACTCTCGTATTAAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 TGGAACCACCGGTTAAGCTCGATGATTTGAGTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy     1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy     2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy     2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149


RESULT 3
US-10-027-880-4
; Sequence 4, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
;  APPLICANT: HOFFMAN, BEATE
;  TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
;  TITLE OF INVENTION:  VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
;  TITLE OF INVENTION:  PROMOTER AND TRANSGENIC PLANTS OBTAINED
;  FILE REFERENCE: CHEP:003US
;  CURRENT APPLICATION NUMBER: US/10/027,880
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: PCT/FR00/01768
;  PRIOR FILING DATE: 2000-06-23
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 4309
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: pBin19 vector
;   OTHER INFORMATION: insert
US-10-027-880-4

  Query Match             100.0%;  Score 2149;  DB 15;  Length 4309;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGATATATAAACAACCG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       30 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGATATATAAACAACCG 89

Qy       61 GCGAAAGTACAAGTTCTACCTTTTTTGGCATGGAACCATGTTTTAGGATTTACTTTGT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       90 GCGAAAGTACAAGTTCTACCTTTTTTGGCATGGAACCATGTTTTAGGATTTACTTTGT 149

Qy      121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      150 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 209

Qy      181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      210 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 269

Qy      241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      270 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 329

Qy      301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      330 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 389

Qy      361 GTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      390 GTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 449

Qy      421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      450 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 509

Qy      481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      510 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 569

Qy      541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      570 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 629

Qy      601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      630 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 689

Qy      661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      690 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 749

Qy      721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      750 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 809

Qy      781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      810 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 869

Qy      841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      870 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 929

Qy      901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      930 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 989

Qy      961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      990 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1049

Qy     1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1050 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1109

Qy     1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1110 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1169

Qy     1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1170 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1229

Qy     1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1230 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1289

Qy     1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1290 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1349

Qy     1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1350 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1409

Qy     1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1410 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1469

Qy     1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1470 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1529

Qy     1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
```

```
Db       1530 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1589

Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Db       1590 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1649

Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Db       1650 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1709

Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Db       1710 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1769

Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Db       1770 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1829

Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Db       1830 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1889

Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Db       1890 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1949

Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Db       1950 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2009

Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Db       2010 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2069

Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Db       2070 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2129

Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149

Db       2130 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2178
```

RESULT 4
US-10-027-880-3
; Sequence 3, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
; FILE REFERENCE: CHEP:003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4413
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-027-880-3

Query Match 100.0%; Score 2149; DB 15; Length 4413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Db        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195

Qy         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Db        196 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255

Qy        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Db        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315

Qy        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Db        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375

Qy        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Db        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435

Qy        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Db        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495

Qy        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Db        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555

Qy        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Db        556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615

Qy        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Db        616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675

Qy        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Db        676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735

Qy        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Db        736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795

Qy        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Db        796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855

Qy        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Db        856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915

Qy        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Db        916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975

Qy        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Db        976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035

Qy        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Db       1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095

Qy        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Db       1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155

Qy       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Db       1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215

Qy       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Db       1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275
```

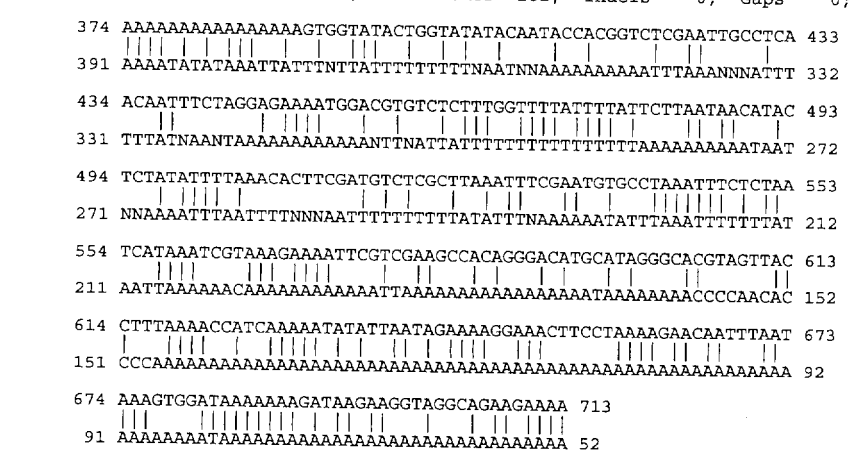
```
Qy       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200
Db       1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335

Qy       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260
Db       1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395

Qy       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320
Db       1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455

Qy       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380
Db       1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515

Qy       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440
Db       1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575

Qy       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
Db       1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635

Qy       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560
Db       1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695

Qy       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
Db       1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755

Qy       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680
Db       1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815

Qy       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740
Db       1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875

Qy       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800
Db       1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935

Qy       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860
Db       1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995

Qy       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920
Db       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055

Qy       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
Db       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115

Qy       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040
Db       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175

Qy       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100
Db       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235

Qy       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
Db       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
```

RESULT 5
US-10-221-714A-254/c
; Sequence 254, Application US/10221714A
; Publication No. US20040048254A1

```
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 254
;   LENGTH: 7008
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-254

  Query Match             3.0%;  Score 63.6;  DB 13;  Length 7008;
  Best Local Similarity   44.1%;  Pred. No. 0.021;
  Matches  264;  Conservative    0;  Mismatches  334;  Indels    0;  Gaps    0;

Qy        116 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 175
Db       4997 TTTCTCATATATAAACCTTCCTTAACAATAATTATTAAAAATTTCTTACTAATCACAAAA 4938

Qy        176 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 235
Db       4937 CTACTATTTAAATAATAATTTAAAATAAAAAAAACCAATATATTTTCATCTCAAATCAAA 4878

Qy        236 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 295
Db       4877 TCAAATATAATCCAACCAACATCTATCAATACTAAAACTAAACTCCATAAAAACAAACAT 4818

Qy        296 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 355
Db       4817 CATATCCATTTCCCTAACACTCAACCAAAAATTTTTATATAATAAAAACTTAATAAATAT 4758

Qy        356 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 415
Db       4757 TTAAATATATTAATAATTAAATAAATAAAAAACCGAATAATAAACCAATAAATAACTACT 4698

Qy        416 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 475
Db       4697 CAACCGAAAACACCGAAAACACAAAAATAAACAAAACATAAACCTTAATCTTCAAAAATA 4638

Qy        476 TTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 535
Db       4637 TATACATTTATCACTTAATTTATATTTTTAACAAATATTTATTTAATATCAACTATTTTC 4578

Qy        536 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 595
Db       4577 CAAACACTTTATTAAATACTACAATAAATAAAACAACTATATTTACTCCCTAAAAATAAA 4518

Qy        596 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 655
Db       4517 AATATTACATAATAATCGTATAAATAACTATAAAATTAAAATTATTTTAAATACAATAAA 4458

Qy        656 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAA 713
Db       4457 AAATTTTATAAAAATCGTACAATAAAAAAAAAATAATCTAAATACAAAAAAAAATAAAA 4400
```

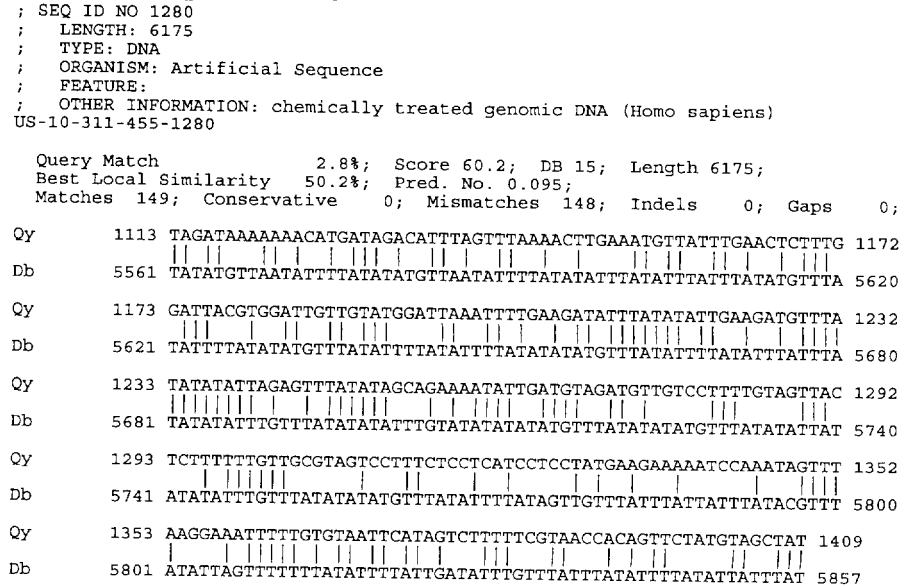
RESULT 6
US-10-021-323-17801/c
; Sequence 17801, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 17801
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(396)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3831-002-Q1-N6-A7
US-10-021-323-17801

Query Match 2.8%; Score 60.6; DB 17; Length 396;
Best Local Similarity 46.8%; Pred. No. 0.021;
Matches 159; Conservative 0; Mismatches 181; Indels 0; Gaps 0;



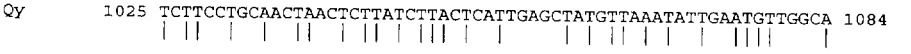
RESULT 7
US-10-311-455-1280
; Sequence 1280, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1280
; LENGTH: 6175
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1280

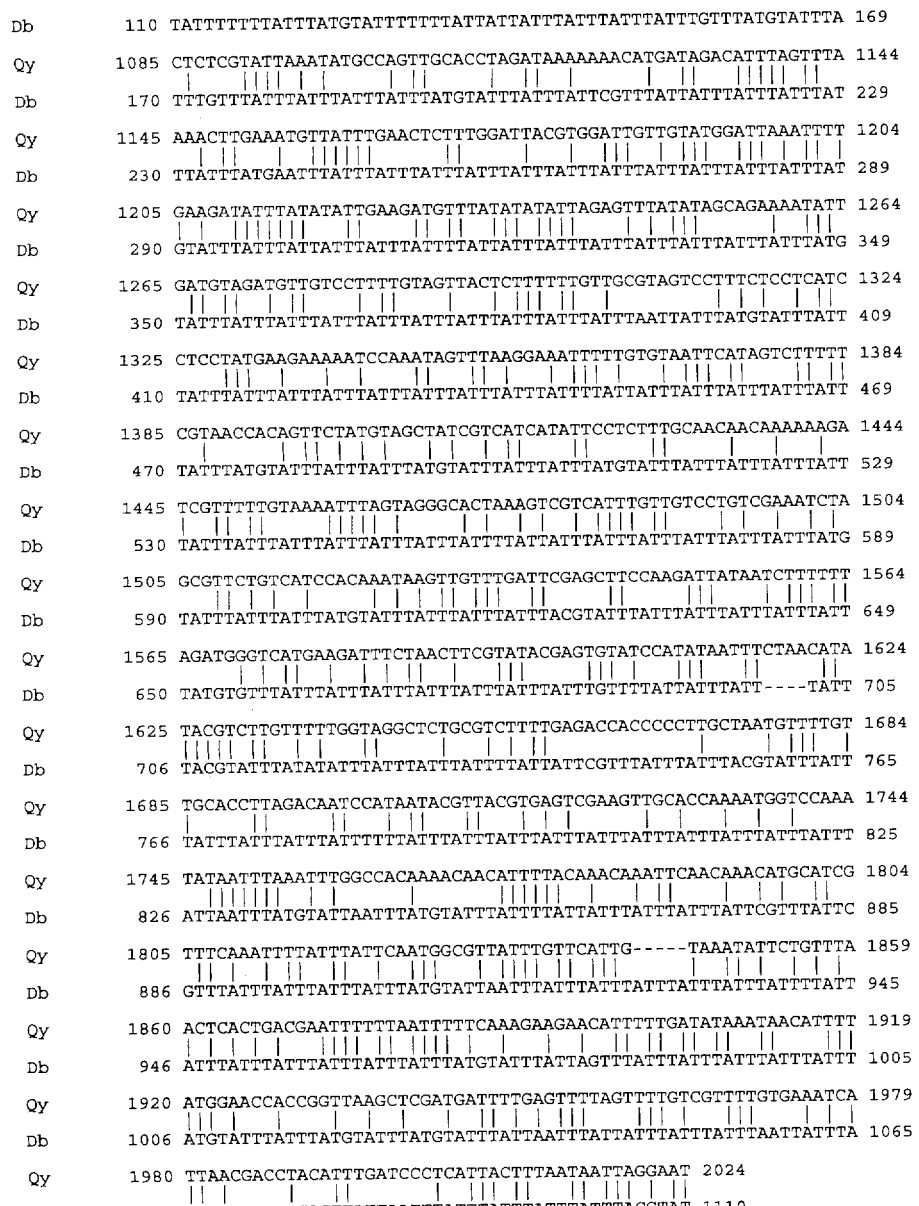
Query Match 2.8%; Score 60.2; DB 15; Length 6175;
Best Local Similarity 50.2%; Pred. No. 0.095;
Matches 149; Conservative 0; Mismatches 148; Indels 0; Gaps 0;



RESULT 8
US-10-311-455-986
; Sequence 986, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 986
; LENGTH: 7597
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-986

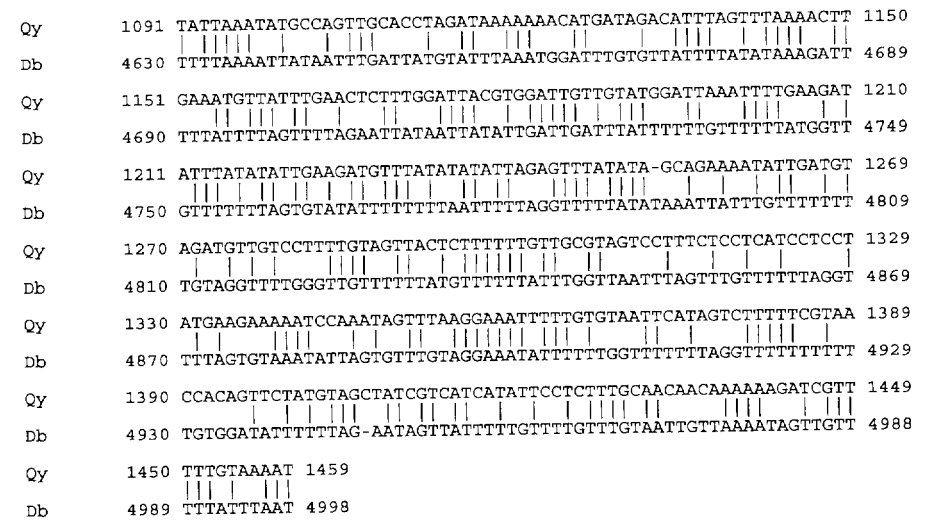
Query Match 2.8%; Score 59.8; DB 15; Length 7597;
Best Local Similarity 42.7%; Pred. No. 0.13;
Matches 429; Conservative 0; Mismatches 567; Indels 9; Gaps 2;





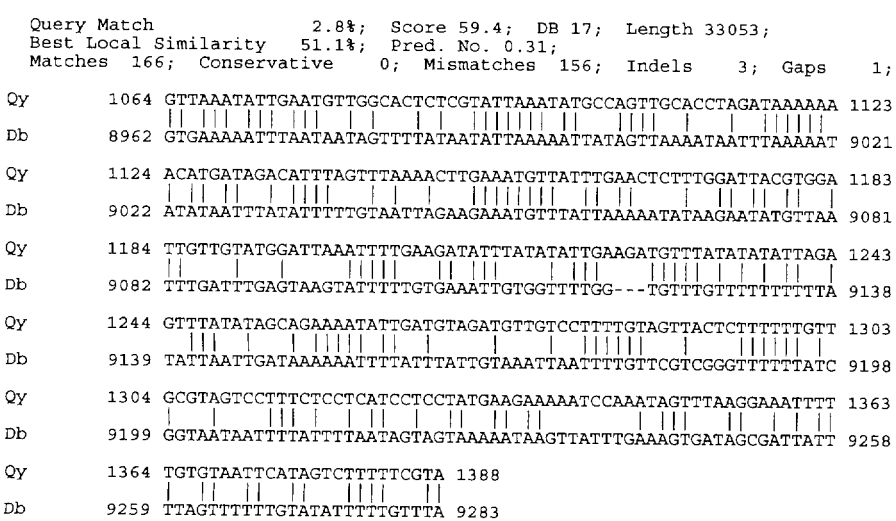
RESULT 9
US-10-311-455-1070
; Sequence 1070, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1070
; LENGTH: 5822
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1070

Query Match 2.8%; Score 59.6; DB 15; Length 5822;
Best Local Similarity 51.1%; Pred. No. 0.12;
Matches 189; Conservative 0; Mismatches 179; Indels 2; Gaps 2;



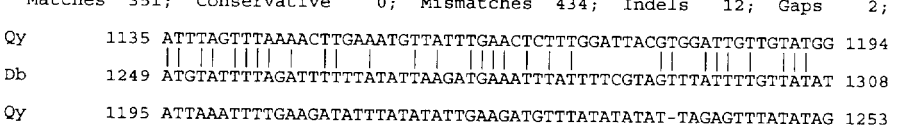
RESULT 10
US-10-433-793-35
; Sequence 35, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 35
; LENGTH: 33053
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-35

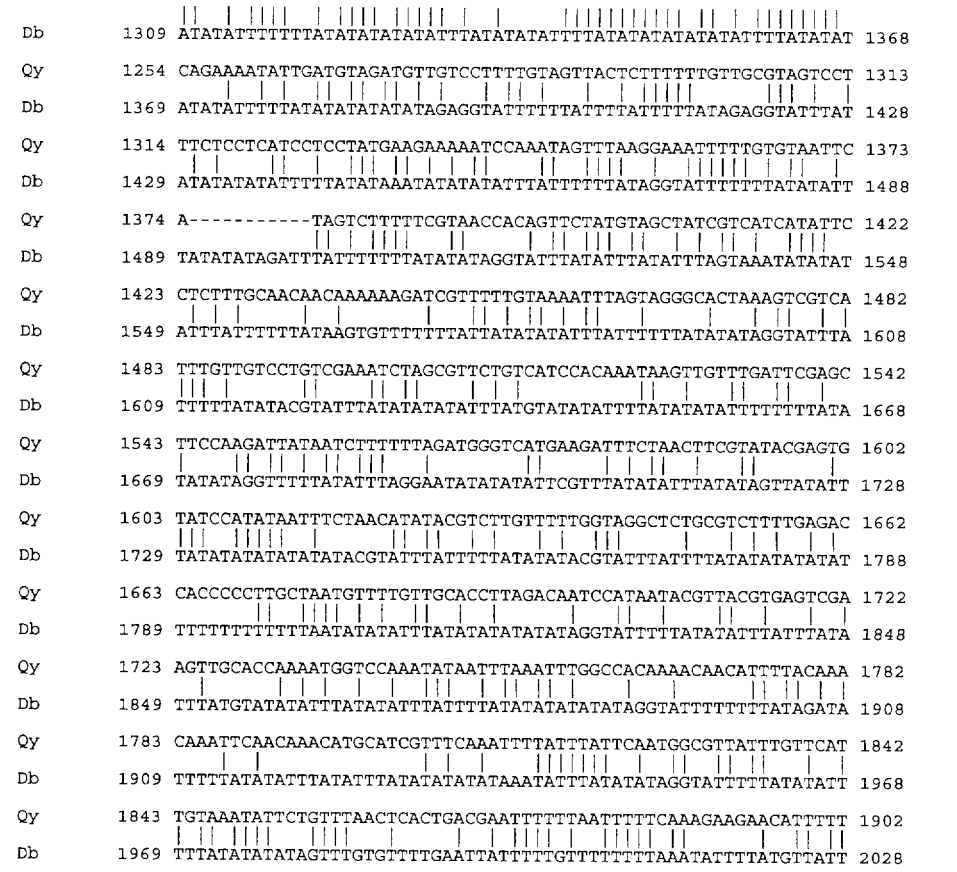
Query Match 2.8%; Score 59.4; DB 17; Length 33053;
Best Local Similarity 51.1%; Pred. No. 0.31;
Matches 166; Conservative 0; Mismatches 156; Indels 3; Gaps 1;



RESULT 11
US-10-221-714A-418
; Sequence 418, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 418
; LENGTH: 5413
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-418

Query Match 2.7%; Score 58.6; DB 13; Length 5413;
Best Local Similarity 44.0%; Pred. No. 0.19;
Matches 351; Conservative 0; Mismatches 434; Indels 12; Gaps 2;





RESULT 12
US-10-433-793-28
; Sequence 28, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 28
; LENGTH: 37515
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-28

Query Match 2.7%; Score 57.6; DB 17; Length 37515;
Best Local Similarity 48.8%; Pred. No. 0.77;

```
  Matches  156;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy       1146 AACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTG 1205
Db       6823 AAGTATAAGTTTTTAGAGATATTTTTAGTTTTTTTTTGTAGTAAAGTTAATAAAATTTTA 6882

Qy       1206 AAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTG 1265
Db       6883 ATAATATTTAAAGGATGTTTATGGTTATTTGGTAAAAAGAATAATTATAATTGATAAAGA 6942

Qy       1266 ATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCATCC 1325
Db       6943 TTGTAGTATTTGATATTTAGAATTAATTATTTTGGATTTTTTTGTTATTTTTAAATAAAA 7002

Qy       1326 TCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTTTTC 1385
Db       7003 TGATTTGTGAGAAATGTATTTTTTTTTAGGTAATATTTTTTTTTTAATTGTATTTATTAT 7062

Qy       1386 GTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAAAAGAT 1445
Db       7063 ATAAATTAAATTAAATTTGAATAATCGTATAAAGTTATTTTTTAAAAAAATAAAATAAGT 7122

Qy       1446 CGTTTTTGTAAAATTTAGTA 1465
Db       7123 AAATATTGTTGTAATTAATA 7142
```

RESULT 13
US-10-311-455-1301
; Sequence 1301, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1301
; LENGTH: 8277
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1301

```
  Query Match             2.6%;  Score 56.8;  DB 15;  Length 8277;
  Best Local Similarity   53.1%;  Pred. No. 0.54;
  Matches  121;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       1096 AATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAAT 1155
Db       6529 AATATACGTGTTAAAATTAGGAAATTAATATGGATAAAATATTATTATTTATTATAGAT 6588

Qy       1156 GTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTA 1215
Db       6589 TTTATTGAGGTTTTTTCGATTATTTTGTTTGTTTTTTGTAGTGAAAATTTTTTTTAGGTT 6648

Qy       1216 TATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGT 1275
Db       6649 TAGGATTTAGTTAAGGATTAATGTTATAGTTTTTAATTTTTTTTAATTTGGATTAGTTTT 6708

Qy       1276 TGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCAT 1323
Db       6709 TTTTTTTTTTTTTTTTTTTTTTTTTGGATACGGAATTTTATTTTGTCGT 6756
```

RESULT 14
US-10-221-714A-461
; Sequence 461, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 461
; LENGTH: 6292
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-461

```
  Query Match             2.6%;  Score 56.6;  DB 13;  Length 6292;
  Best Local Similarity   44.5%;  Pred. No. 0.52;
  Matches  224;  Conservative    0;  Mismatches  279;  Indels    0;  Gaps    0;

Qy       1136 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1195
Db       3907 TTTTGGTTATATTTTTATTTTTTTTTATTTGTTATTTTTTTTTTTTTTTTCGTTTATT 3966

Qy       1196 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1255
Db       3967 TAAATATTTTTTAGTTATTGTTTTTTTTTTTTATATTTTTTTTTTTGTTTATCGTTTTTT 4026

Qy       1256 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTT 1315
Db       4027 AGTAAAATTTTTTTTTTTTTCGGTTCGTTATTTTTTTTATTTTTTTTTTATTGGTTATTT 4086

Qy       1316 CTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCAT 1375
Db       4087 TTTTTTTTTTTTTTATTTATTTAAAAGTTTTTTTTTTTATTGTTTTTTTATAAAATTTTT 4146

Qy       1376 AGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAA 1435
Db       4147 TTTTTTTTTTGTTCGTTATTTTTTTTTTTTTTTTTTTTTTTTGTATTTTTTTTTTTTTTT 4206

Qy       1436 CAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGT 1495
Db       4207 TATTTGTTATTTTTTTTTTTTTTTTTTATTATTTAAAATTTTTTTATTTATAGTTTTTTTT 4266

Qy       1496 CGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATA 1555
Db       4267 TTTTTTTTTTTTTTTTTATTATATTTTTGTAAATTTTTTTTTTTTTTTGTTTATTTTCGTT 4326

Qy       1556 ATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1615
Db       4327 TTTTTTTTACGATTTTTTTTTATTTTTTTTTATTTATTTAAAAATTTTTTTTTTTATTATT 4386
```

```
Qy       1616 TCTAACATATACGTCTTGTTTTT 1638
Db       4387 TTTTGTGAAATTTTTTTTTTTTT 4409

RESULT 15
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             2.6%;  Score 56.6;  DB 15;  Length 3673778;
  Best Local Similarity   46.1%;  Pred. No. 11;
  Matches  298;  Conservative    0;  Mismatches  344;  Indels    5;  Gaps    3;

Qy        778 ATTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATA 837
Db    1714728 ATTTTATATGTATATTTTATATATATATTTTATATGTATATTTTATATATATATTTTATA 1714787

Qy        838 TCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATT 897
Db    1714788 TATATTTTATATATATATT-TTATATATATATTATTTATATTTTATATATATAGTATATA 1714846

Qy        898 CGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTAT---GTTT 954
Db    1714847 TATTTTATATGTATATATTTTATATATATTTTATATATATATATTTTATATATATATTTT 1714906

Qy        955 CTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTT 1014
Db    1714907 ATATATATTTTATATATATATTTGTATATATATATTTTGTATATATATATTTTATATAT 1714966

Qy       1015 GAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTG 1074
Db    1714967 ATATTTTATATATATATATATTTTATATATATATATATTTTATATATATTTTATATATAT 1715026

Qy       1075 AATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGAC 1134
Db    1715027 ATATATTTTATATATATATTTTATATATATATATATTTTATATATATATTTTATATATAT 1715086

Qy       1135 ATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGG 1194
Db    1715087 ATATATTTTATATATATATATTTTATATATATATTTTATATATATATTTTATATATATAT 1715146

Qy       1195 ATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGC 1254
Db    1715147 TTTATATATATATTTTATATATATATTTTATATATATATATTTTATATATATATTTTATA 1715206

Qy       1255 AGAAAATATTG-ATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCT 1313
Db    1715207 TATATATATTTTATATATATATATTTTATATATATATTTTATATATATATTTTATATATATA 1715266

Qy       1314 TTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTC 1373
Db    1715267 TTTTATATATATATTTTATATATATATTTTATATATATATTTTATATATATATTTTATAT 1715326

Qy       1374 ATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATAT 1420
Db    1715327 ATATATTTTATATATATATATTTTATATGTATATTTTATATATATAT 1715373

Search completed: July 31, 2004, 10:49:29
Job time : 993 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 04:03:48 ; Search time 5558 Seconds
(without alignments)
11546.208 Million cell updates/sec

Title: US-10-027-880-1
Perfect score: 2149
Sequence: 1 gtcgaattgtgatatattgt.....actaattgctaattaatcag 2149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	355	16.5	355	29	CC793681	CC793681 SALK_0173
c	2	159	7.4	833	28	BZ457686	BZ457686 BONRW80TR
	3	114.4	5.3	703	28	BZ470713	BZ470713 BONNA88TR
c	4	98.8	4.6	690	28	BH959481	BH959481 odd91g04.
c	5	87.2	4.1	92	29	BX285678	BX285678 Arabidops
c	6	85.2	4.0	1200	13	BX437758	BX437758 BX437758
	7	80	3.7	1101	29	CNS0039G	AL063921 Drosophil
c	8	78.6	3.7	646	28	BH963752	BH963752 odf98b01.
c	9	76	3.5	1201	9	AL565455	AL565455 AL565455
c	10	73.6	3.4	1101	29	CNS00EVL	AL069706 Drosophil
c	11	73	3.4	928	29	CNS00DKY	AL071865 Drosophil
c	12	71.4	3.3	379	28	BH757746	BH757746 SALK_0081
	13	70.8	3.3	604	28	AQ964121	AQ964121 LERGT42TR
c	14	70.8	3.3	697	28	AQ964120	AQ964120 LERGT42TF
	15	70.6	3.3	1200	13	BX437758	BX437758 BX437758
	16	69.8	3.2	1896	29	CG753083	CG753083 P048-1-C0
c	17	69.4	3.2	1101	29	CNS00EO7	AL069440 Drosophil
	18	69.4	3.2	1200	13	BX415878	BX415878 BX415878
c	19	69.2	3.2	726	28	B26447	B26447 F2K20TF IGF
c	20	68.4	3.2	1200	13	BX415878	BX415878 BX415878
c	21	67.8	3.2	1101	29	CNS0039G	AL063921 Drosophil
c	22	67.8	3.2	1201	13	BX439779	BX439779 BX439779
	23	67.2	3.1	961	14	CK230311	CK230311 ILLUMIGEN
c	24	66.2	3.1	83	29	AL763344	AL763344 Arabidops
c	25	66	3.1	1201	9	AL536104	AL536104 AL536104
	26	66	3.1	1201	13	BX360615	BX360615 BX360615
c	27	65.8	3.1	956	13	BX414969	BX414969 BX414969
c	28	65.6	3.1	1165	13	BX338369	BX338369 BX338369
c	29	65.2	3.0	1201	13	BX443774	BX443774 BX443774
	30	65	3.0	1201	13	BX462207	BX462207 BX462207
	31	64.8	3.0	994	13	BX414650	BX414650 BX414650
c	32	64.6	3.0	701	28	AQ250248	AQ250248 T3F21-Sp6
c	33	64.6	3.0	1056	13	BX415058	BX415058 BX415058
	34	64.6	3.0	1206	28	BZ695529	BZ695529 SP__Ba006
	35	64.4	3.0	1101	29	CNS0039R	AL063932 Drosophil
c	36	64.2	3.0	1101	29	CNS002M0	AL062875 Drosophil
	37	64.2	3.0	1101	29	CNS00EVL	AL069706 Drosophil
	38	64.2	3.0	1392	29	CG757503	CG757503 P052-4-C0
	39	64	3.0	1167	29	CNS07360	AL427102 clone BA0
c	40	64	3.0	1201	13	BX461824	BX461824 BX461824
	41	63.8	3.0	1200	13	BX437739	BX437739 BX437739
	42	63.8	3.0	1201	9	AL536104	AL536104 AL536104
	43	63.4	3.0	1029	29	CNS01ZGM	AL174271 Tetraodon
	44	63.4	3.0	1098	13	BX377526	BX377526 BX377526
c	45	63.2	2.9	1101	29	CNS016LI	AL106896 Drosophil

ALIGNMENTS

RESULT 1
CC793681
LOCUS CC793681 355 bp DNA linear GSS 01-JUL-2003
DEFINITION SALK_017399.55.00.x Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_017399.55.00.x, genomic survey sequence.
ACCESSION CC793681
VERSION CC793681.1 GI:32388904
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 355)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

```
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .355
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_017399.55.00.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             16.5%;  Score 355;  DB 29;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-51;
  Matches  355;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1790 AACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAAT 1849
Db          1 AACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAAT 60

Qy       1850 ATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAA 1909
Db         61 ATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAA 120

Qy       1910 ATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTT 1969
Db        121 ATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTT 180

Qy       1970 TGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAAC 2029
Db        181 TGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAAC 240

Qy       2030 ATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATG 2089
Db        241 ATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATG 300

Qy       2090 ACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTA 2144
Db        301 ACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTA 355


RESULT 2
BZ457686/c
LOCUS       BZ457686               833 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONRW80TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONRW80,
            genomic survey sequence.
ACCESSION   BZ457686
VERSION     BZ457686.1  GI:26737311
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BONRW80TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .833
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONRW80"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             7.4%;  Score 159;  DB 28;  Length 833;
  Best Local Similarity   62.6%;  Pred. No. 1.2e-17;
  Matches  420;  Conservative    0;  Mismatches  200;  Indels   51;  Gaps    9;

Qy        584 CACAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAA-TATATTAATAG 642
Db        818 CAAAGGGACATGCATGGGGCACGTAGTTACTTTTAGACCATAAAAAAATTAAACTAATAG 759

Qy        643 AAAAGGAAACTTCCTAAAAGAACAATT-------TAATAAAGTGGATAAAAAAAGATAAG 695
Db        758 AGAAGGAAACTTCCTAAATAAAAGATAAATAAAATAAAAAAGGAAAGAAAAAACAAAGAA 699

Qy        696 AAGGTAGGCAGAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGC 755
Db        698 GGTATAGGCAAAAGAAAACGTATAGCCGCGACCCGTAACAAAGGACGTCCTGACCACTGC 639

Qy        756 GGAGACGGCGAGACGCTGACTGATTTTTTCT---TTTTCTTTTCCTAAAGAACGTTGTTT 812
Db        638 GGAGACGGCGAGACGCTGACTAAATTTTTTTATCTTTTGTTTTCCTAAAGAACGTTGTTT 579

Qy        813 CGTGCTTACAAGGGTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGA 872
Db        578 CGTGCTCACAAGGGTCAA-----------------GTGTCAACCATATCAAGATATAAAA 536

Qy        873 TCCCCTCTTGTGCTTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAA 932
Db        535 CCCCTTCTTGTAGTTTGTCTTTACTCGTGATAAAT-CTAAGTTTGAAATTAGTTCTAATT 477

Qy        933 TATATATGTCCTACCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATAT 992
Db        476 TCAAAATTATATATGCAT----------------GTGACTCTATTTACCTCGGGTCCCTT 433

Qy        993 GTGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGC---AACTAACTCTTATCTT 1049
Db        432 GTGAACATGACGTCTGAAATTTGAAGAGCTAGTTGTTCTACCTGCAACAAATTATTTCTT 373

Qy       1050 ACTCATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCG--TATTAAATATGCCAGTT 1107
Db        372 AAGCATTGTTCAGTGCTAAGTACTGAAAGCTGACAGTCTTGTATATATAATATGGCATCT 313

Qy       1108 GCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACT 1167
Db        312 GCACCTACCAAAACCATTGATAATTATTTTAAACTTATTTTTTGCTAACAAGGATTGTC- 254

Qy       1168 CTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGAT 1227
Db        253 GTATGAATTTTAAAGATGTATCGTATACGTAACTAATAAACGTATATAGATATTCCAGAT 194

Qy       1228 GTTTATATATA 1238
Db        193 CCATATATTTA 183


RESULT 3
BZ470713
LOCUS       BZ470713               703 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONNA88TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONNA88,
            genomic survey sequence.
```

```
ACCESSION   BZ470713
VERSION     BZ470713.1  GI:26768062
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BONNA88TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONNA88"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             5.3%;  Score 114.4;  DB 28;  Length 703;
  Best Local Similarity   77.6%;  Pred. No. 5.4e-10;
  Matches  166;  Conservative    0;  Mismatches   41;  Indels    7;  Gaps    2;

Qy        584 CACAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAA-TATATTAATAG 642
              || |||||||||||| |||||||||||||| |||| | |    ||||| || | ||||||
Db        490 CAAAGGGACATGCATGGGGCACGTAGTTACTTTTAGACCATAAAAAAATTAAACTAATAG 549

Qy        643 AAAAGGAAACTTCCTAAAAGAAC------AATTTAATAAAGTGGATAAAAAAAGATAAGA 696
              | ||||||||||||||||  ||       ||  ||| ||||   | ||||||  |  |
Db        550 AGAAGGAAACTTCCTAAATAAAGATAAATAAAATAAAAAAGGAAAGAAAAAACAAAGAAG 609

Qy        697 AGGTAGGCAGAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCG 756
              ||||||  |||||||||||||  ||||||||  ||||||||  |||||||  ||||||||||
Db        610 GTATAGGCAAAAGAAAACGTATAGCCGCGACCCGTAACAAAGGACGTCCTGACCACTGCG 669

Qy        757 GAGACGGCGAGACGCTGACTGATTTTTTCTTTTT 790
              |||||||||||||||||| | ||||| | | |
Db        670 GAGACGGCGAGACGCTGACTAAATTTTTTTATCT 703


RESULT 4
BH959481/c
LOCUS       BH959481                690 bp    DNA     linear   GSS 01-OCT-2002
DEFINITION  odd91g04.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH959481
VERSION     BH959481.1  GI:23440708
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: odd91  row: g  column: 04
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 16
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
ORIGIN

  Query Match             4.6%;  Score 98.8;  DB 28;  Length 690;
  Best Local Similarity   84.2%;  Pred. No. 2.5e-07;
  Matches  123;  Conservative    0;  Mismatches   22;  Indels    1;  Gaps    1;

Qy       2005 TTACTTTAATAATTAGGAATCAAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGG 2064
              ||  ||   | ||||  ||  ||||| |||||| ||||||| ||| |||||| |||||||
Db        551 TTTATTACTTTATTACCAAGAAAACAAGATGATAAAGTTCATCAACGACGTCCCTTATGG 492

Qy       2065 CTATTAAGAGTCAGACGCAAGGATGACCGGGGTCATTAAGACGTCTTATAT-TCAACCAT 2123
              |||||||||||||| |||||||||||| |||||||||||| ||||||||   ||||||
Db        491 CTATTAAGAGTCAGATGCAAGGATGACCAGGGTCATTAAGACATCTTATATCGAAACCAT 432

Qy       2124 TACTCCACTAATTGCTAATTAATCAG 2149
              ||||||||||||| ||||||| ||||
Db        431 TACTCCACTAATTACTAATTATTCAG 406


RESULT 5
BX285678/c
LOCUS       BX285678                92 bp    DNA     linear   GSS 07-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-384F07-017271,
            genomic survey sequence.
ACCESSION   BX285678
VERSION     BX285678.1  GI:28884674
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 92)
  AUTHORS   Rosso,M., Strizhov,N., Li,Y. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2003) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion within the locus defined by clone T32E8. The
```

sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.

FEATURES Location/Qualifiers
source 1. .92
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-384F07-017271"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thalaina nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

ORIGIN

Query Match 4.1%; Score 87.2; DB 29; Length 92;
Best Local Similarity 96.7%; Pred. No. 7.6e-05;
Matches 89; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1873 TTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACCACCGG 1932
Db 92 TTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACCCCCGG 33

Qy 1933 TTAAGCTCGATGATTTTGAGTTTTAGTTTTGT 1964
Db 32 TTAAGGTCGAAGATTTTGAGTTTTAGTTTTGT 1

RESULT 6
BX437758/c
LOCUS BX437758 1200 bp mRNA linear EST 15-MAY-2003
DEFINITION BX437758 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YB01 5-PRIME, mRNA sequence.
ACCESSION BX437758
VERSION BX437758.1 GI:30773605
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0CAP008CA01QP1.
FEATURES Location/Qualifiers
source 1. .1200
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP008YB01"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 4.0%; Score 85.2; DB 13; Length 1200;
Best Local Similarity 31.2%; Pred. No. 4e-05;
Matches 278; Conservative 174; Mismatches 425; Indels 15; Gaps 2;

Qy 780 TTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATC 839
Db 1200 TTWTTATYTTAWTTTTTTTTYTTWTYTTYWCTTTTTTYTTTTTATTTTCTTWTTHTMTTC 1141

Qy 840 CAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCG 899
Db 1140 TTTTTTTTTTTTTTTTTTATWTTTTYTTTTTTTTYTAAWATTTTTCTTTTTTTAWAAAAA 1081

Qy 900 TGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACT 959
Db 1080 AAAAAAAAAWTTTATCTCTHMTATWTCTTTYTTHTWATAATTYWATHYTATTTWTCTATA 1021

Qy 960 GACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGA 1019
Db 1020 WTAWATAWATAAAAHWWWTTTTYTWAWAAAAAAAWWTTTTTWWYATATTAAAAAAWATTW 961

Qy 1020 GTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGT 1079
Db 960 TTWTCTTYTTCTTTTTWTWTTYTTTYWWTWTWAAATWTTTTTTATAWAAWWTTTTTATTT 901

Qy 1080 TGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTA 1139
Db 900 TTTTTTCMWYYTTYTTTTATTTTATTTTTYTMTTATATATAAAWATATATTWYTYWTTTT 841

Qy 1140 GTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAA 1199
Db 840 TTTTTTWTTWAWATMTTTYTCTWWWAATTTTTTTTAWA-----AWWWWTTTTTMHATCTT 786

Qy 1200 ATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAA 1259
Db 785 WTWWAAAAAATAWWTATAAAAAAHTCYTCYTYATAWWTWTAAWATCYYTCTMATAAAAAA 726

Qy 1260 ATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCC 1319
Db 725 AWTTYCTTCTCYTTTTCTWWTWTTTWAWAWATYYMYTTCCTWTWTAWWAWAAWWMTATMT 666

Qy 1320 TCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTC 1379
Db 665 AYYTCHYWWWAAACAAWAAAAAAAAAWAAAAAAAWAWAWMWHCYTTWAWAAAAAAAAWATC 606

Qy 1380 TTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAA 1439
Db 605 CTYMWWATAAAMMAAAAWATMWTCTSTTTTTWTWAWWAAAAATATAWATWWWTAAAAAAA 546

Qy 1440 AAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAA 1499
Db 545 AAAAAAWAAWAWWTAYCTCCCWWWTTAAAAAAAATCYTYCYTTWWWAWWTTTTTTTYWWA 486

Qy 1500 ATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCT 1559
Db 485 AWATTYTTTTTTMTWAWAAAAAAAAAAAATTTT----------TYYMMAAAAAAAAMYMH 436

Qy 1560 TTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTA 1619
Db 435 WWWTWTTWAWWWWHTTYAHAACHYMYTCYCTYWWAAAAAAWATTCTHTWTMTHHHHWWWW 376

Qy 1620 ACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTT 1671
Db 375 WATAAAAWAWAWWTAWTHTCCYCCYMTATYTTTTTWAAAAAAAAAMHYYYTT 324

RESULT 7
CNS0039G
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit

```
          fly), genomic survey sequence.
ACCESSION AL063921
VERSION   AL063921.1  GI:4941778
KEYWORDS  GSS.
SOURCE    Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             3.7%;  Score 80;  DB 29;  Length 1101;
  Best Local Similarity   17.5%;  Pred. No. 0.00032;
  Matches  121;  Conservative  286;  Mismatches  283;  Indels    0;  Gaps    0;

Qy         12 ATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCGGCGAAAGTACA 71
              |||:|:: ::     | | :||| || |       : |:|:|:|| ||      ||:  | :
Db        398 ATAWAWWWWWTTTTTTTTAWAAAWAAAATAATTWWAAWAWAAAAAATTWWAAAAWAAAAW 457

Qy         72 AGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAAT 131
              | |  :    :|:       |   ||  :| ||||||:    :|   |  :  ::  | ||:
Db        458 AWTAWWTTTAWTWAAAAAAAAAAAATTWTTTTTTTWTTTAWTTWATAWWTTWWWTTAAAW 517

Qy        132 CTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACAAGTTCTACCAA 191
                    |     : ||:  |:||  | :|::|::|    :||:|   |: :  :::: :
Db        518 AAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWATAAAMCMAAWYYHTYTT 577

Qy        192 AGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAATCTTATGATTT 251
              :::   : :|: :: :  : : : :  : : ::    ::::  :: :: :::: :|
Db        578 YTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAWWHWTTYTA 637

Qy        252 CCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTGTCATTAACTTC 311
               :::::|| :: : ::::: : : | ::| :|:: :|:: ||:: ::: :||   :||:
Db        638 AYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYHYYYYMYCAMMCMCTHT 697

Qy        312 CCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTG 371
              |::|::: :::  :: : ::::|:       ::| :::: | | ::  :: :  :|::||
Db        698 CHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWMYCTACTYHYHHHHHYHWAYHTTW 757

Qy        372 TGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTCTCGAATTGCCT 431
              : :|:|::::::|: |||     :   :  ::| : :::| :::  :  | : |: | :
Db        758 YAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYHYYMYTCCYMCTYHCWH 817

Qy        432 CAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATTCTTAATAACAT 491
                 : : | | : :   : : : :  :: : ::| |::  ::|: : ::::: ||: :| : :
Db        818 YYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWCWWWWHATTWTWATHCW 877

Qy        492 ACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCCTAAATTTCTCT 551
              ||   ::::: ::::::::  : :: |: | |::::: | : ::| : :  ::::: ::::
Db        878 ACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHTMYHMTCHWMWHMHWHW 937

Qy        552 AATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAGGGCACGTAGTT 611
              ::: :  ::|  | :: :  : ::  :  : ::::  ::   |:|:  : : :
Db        938 HHWWMATWMTTMTTMMMMCCMMHHHCHMYHMMHMYMYCCHYYCTCHTHATTHYHYMCTCY 997

Qy        612 ACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAAAGAACAATTTA 671
              ::||::  :  ::: : |:|  |||::::: ::|:   :: ::  :| |: |   :::|:
Db        998 HYCTWHTYWTAYWWAWTAHAMTTATWWWWMHWWAHWATWWWWWWWWATAWWACTCHHWTW 1057

Qy        672 ATAAAGTGGATAAAAAAAGATAAGAAGGTA 701
               : :  :  ::  ::::| : |: ::  :|
Db       1058 YHTHCTWYYHHTYHMWWAWWMAWHWHHMYA 1087


RESULT 8
BH963752/c
LOCUS       BH963752                646 bp    DNA     linear   GSS 01-OCT-2002
DEFINITION  odf98b01.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH963752
VERSION     BH963752.1  GI:23444978
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: odf98  row: b  column: 01
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 64
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
ORIGIN

  Query Match             3.7%;  Score 78.6;  DB 28;  Length 646;
  Best Local Similarity   61.9%;  Pred. No. 0.00076;
  Matches  300;  Conservative    0;  Mismatches  109;  Indels   76;  Gaps    8;

Qy        596 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 655
              |||  |||||||||||||||||| | ||||| |||||    |    |  ||||
Db        646 CATGTGGCACGTAGTTACCTTTTAGACCATTAAAAACTACTCCGAAATAAAG-------- 595

Qy        656 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACG 715
              ||||  | |||| |||  |  | |  |||||  ||||||||||||| |||||||||
Db        594 --AAAAATAAAATTAAATGGAAAGAAAGGAAAAAACAAAGAAGGTAGGCAAAAGAAAACG 537
```

```
Qy        716 TATGGCCGCGACTC--GTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTG 773
              ||| |||||||| |    |  ||||||||||||||||||||||||||| ||||||
Db        536 TATAGCCGCGACCCCGTAACAAAGGGACGTCCCGACCACTGCGGAGATGGCGAG------ 483

Qy        774 ACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAAC 833
                                      ||||| |||||||| ||||||||||
Db        482 ------------------------------ACGTTATTTCGTGCTCACAAGGGTCAAGTG 453

Qy        834 CATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTT 893
              |  |  |  |   | |   || |||   | | |     |  ||| ||||||| ||||||||
Db        452 CGAACGCTCTCTCTATTGATAATATATAAAATCCCCGG--CCCATCTTGTGGTTTGTCTT 395

Qy        894 TATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTT 953
              |||||||| |||| |   |  | |||||||||||| ||||||| |    |  |||||
Db        394 TATTCGTGTTATA-ATCGAAGTTGTACTTAGTTCTAAATTATATATATATGTC--ATGTT 338

Qy        954 TCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATT 1013
              ||||||  |||| ||||||                        |||| |||||||  |  |||||
Db        337 TCTACTTGCCTCGGTCCCT------------------TGTGATAATGACGTCTGAAATT 297

Qy       1014 TGAAGAG----TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAA 1069
              |||||||    ||  || |||| |||||||||  ||  ||||  ||||||  |    |  |
Db        296 TGAAGAGCTATTTGTTCGTCCTCCAACTAACTATTTCCTTAAACATTGATCAGCGCTTTT 237

Qy       1070 TATTG 1074
              | |||
Db        236 TTTTG 232
```

RESULT 9
AL565455/c
LOCUS AL565455 1201 bp mRNA linear EST 12-MAY-2003
DEFINITION AL565455 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF005YO18 3-PRIME, mRNA sequence.
ACCESSION AL565455
VERSION AL565455.2 GI:30549492
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:12916848.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9232.f For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF005BH09NP1&cluster=9232.f. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF005BH09NP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF005YO18"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 3.5%; Score 76; DB 9; Length 1201;
Best Local Similarity 33.4%; Pred. No. 0.0015;
Matches 269; Conservative 127; Mismatches 402; Indels 8; Gaps 2;

```
Qy        860 ATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGTGATATATAATCTAACTTAA 919
              ||:||  ||  : ||    :  | | :|    :|  :|  : ||:|||:   : | :||
Db       1201 ATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTATWTATWTAT 1142

Qy        920 ATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTGACCTCAGTCCCTAGTTAGC 979
              ||:  |:   : ||:|||:| |  ||  ||| |:| |:   :  |   |   |:  |:
Db       1141 ATWTATWTATWTATWTATWTAT-WTATWTATATWTATWTATWTATWTATWTATWTATWTA 1083

Qy        980 TATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAA 1039
              |||:|  :  ||:|  |   | :|         | ||:|  :  | :|  :   |     | : |
Db       1082 TATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTATWTATWTA 1023

Qy       1040 CTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCGTATTAAATA 1099
                :||| |   |  :|     :  |: ||||: | :  |     | |    :|| :||
Db       1022 TATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTATWTATWTA 963

Qy       1100 TGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTA 1159
              |     |:           : | :   :  ||| : ||:|| :|  :  | :|  :   :||
Db        962 TATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTATWTATWTA 903

Qy       1160 TTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTATATA 1219
              | |  | :  |:   :||  |  :|  :| ||| ||: | :| |  | :|||:|||:||
Db        902 TATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTATWTATWTA 843

Qy       1220 TTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGTTGTC 1279
              |       |:|||:|||:|     |:|||:||    | :|||: || || :| |:
Db        842 TA-------TWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTAT 790

Qy       1280 CTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAA 1339
              | |:    |:  | | |:|  :|   ||  :  |:|  |  : ||     |||  | : |
Db        789 WTATWTATATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTAT 730

Qy       1340 ATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCTTTTTCGTAACCACAGTTCT 1399
              :|  : ||| :|  :   :  |:|  :   :|  :||   :| |:  |:   :  | :| |
Db        729 WTATWTATATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTAT 670

Qy       1400 ATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAAAAGATCGTTTTTGTAAAAT 1459
              :| |:  |||    :|  :  |    :|    | :   | : | :|  :| |:   : ||
Db        669 WTATWTATATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTATWTAT 610

Qy       1460 TTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCA 1519
              :|| :    :  :  |    :| | |  :  | |  | :| |:   :| | |  :
Db        609 WTATWTATATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTAT 550

Qy       1520 CAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTTTTTTTAGATGGGTCATGAAG 1579
                | :  | | |    |  |  |     | :|  : || |:| |:   :   : || |
Db        549 WTATWTATATWTATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTAT 490

Qy       1580 ATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAACATATACGTCTTGTTTTTTG 1639
              :| | ||  |:  |:||     | |||  ||:|| :|    :  ||:||  | |:  |:|
Db        489 WTATWTATATWTATWTATWTATWTATWTATWTATATWTATWTATWTATWTATWTATWTAT 430

Qy       1640 GTAGGCTCTGCGTCTTTTGAGACCAC 1665
                |||| | |    | ||  |  |||
Db        429 ATAGGATATATACCCTTCTACTTCAC 404
```

RESULT 10
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)

```
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             3.4%;  Score 73.6;  DB 29;  Length 1101;
  Best Local Similarity   33.8%;  Pred. No. 0.004;
  Matches  175;  Conservative   96;  Mismatches  245;  Indels    2;  Gaps    1;

Qy        786 TTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCCAATTG 845
Db        979 TATAWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTT 920

Qy        846 TTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGTGATAT 905
Db        919 TTWTTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATAT 860

Qy        906 ATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTGACCTC 965
Db        859 TAWATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTW 800

Qy        966 AGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAGTTCCT 1025
Db        799 TTWTATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTW 740

Qy       1026 CTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCAC 1085
Db        739 TTTTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAAT 680

Qy       1086 TCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAACATGATAGACATTTAGTTTAA 1145
Db        679 TATTTAATTAWWWAATWAWTTATTTAATAAWTAATTAA--ATTATTAAAAATTTTATWAA 622

Qy       1146 AACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTG 1205
Db        621 AAAATATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWT 562

Qy       1206 AAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTG 1265
Db        561 ATTWWTTTAAATTATWTTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWT 502

Qy       1266 ATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTT 1303
```

```
Db        501 WTTTTTWATTWATAAWTTTWTTTWWAWATTWWKKKKKK 464


RESULT 11
CNS00DKY/c
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             3.4%;  Score 73;  DB 29;  Length 928;
  Best Local Similarity   33.4%;  Pred. No. 0.0056;
  Matches  136;  Conservative   84;  Mismatches  187;  Indels    0;  Gaps    0;

Qy         21 AAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCGGCGAAAGTACAAGTTCTACC 80
Db        897 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWWYTCMAAAMAMAMMMCVMMMMM 838

Qy         81 TTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAATCTTTCATTT 140
Db        837 MMMMMMMMMMTTMWWHHHWHHHHHHWTTTTTTWTTTTTTTTTTTTTTWWWWWWWWWWWT 778

Qy        141 CTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGA 200
Db        777 TTWWWWTWWWWWWWWWWTNNNWNMCHWAWATWWWWWWATATTTWTATWTWAWATAWAAAA 718

Qy        201 GTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAATCTTATGATTTCCCTTTTCT 260
Db        717 AAAAAAAAAAWTAAATTTTTWTTTTAWTTWTAAAATTTTTTATTWTTTTTTWTTTTWTTT 658

Qy        261 TCACGATATATACTGATATTGATATGCACCCATTTGTTTGTCATTAACTTCCCACTCTAT 320
Db        657 WTATATAAWAWTAWAWAAWAWATATTATATWAWATATTTWWTATTWAWTATAAAATWWTA 598

Qy        321 ACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAA 380
```

```
                :  | | :   :  |:    || |: :| |:  | || |:    ||: || | |||||::
Db        597 WWTTTAAWTAAWATAWAATTAAAAWTWAAAWAAAAAAAAAWAAAATAWTTTWTWAAAAAWW 538

Qy        381 AAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTCTCGAATT 427
              |||||:|   |  | |: |  || | |   ||  |   | |  | :|
Db        537 AAAAAWATTWTTTTTTWWTMATAAAAAATTTAWTAWCATATTTATWT 491


RESULT 12
BH757746/c
LOCUS       BH757746                379 bp    DNA     linear   GSS 01-MAR-2002
DEFINITION  SALK_008127.24.55.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_008127.24.55.x, genomic
            survey sequence.
ACCESSION   BH757746
VERSION     BH757746.1  GI:19042094
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_008127.24.55.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             3.3%;  Score 71.4;  DB 28;  Length 379;
  Best Local Similarity   98.6%;  Pred. No. 0.017;
  Matches   72;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 154

Qy         61 GCGAAAGTACAAG 73
              |||||||||||| |
Db        153 GCGAAAGTACATG 141


RESULT 13
AQ964121
LOCUS       AQ964121                604 bp    DNA     linear   GSS 28-JAN-2000
DEFINITION  LERGT42TRB LERG Arabidopsis thaliana genomic clone LERGT42, genomic
            survey sequence.
ACCESSION   AQ964121
VERSION     AQ964121.1  GI:6791710
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="LANDSBERG ERECTA"
                     /db_xref="taxon:3702"
                     /clone="LERGT42"
                     /clone_lib="LERG"
                     /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                     sheared to 0.4-0.7 Kbp before ligation."
ORIGIN

  Query Match             3.3%;  Score 70.8;  DB 28;  Length 604;
  Best Local Similarity   86.7%;  Pred. No. 0.017;
  Matches   78;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy         88 GGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAAT 147
              || |||| ||||| ||||||| |||||||||||||||| ||  ||||||||||| |||||
Db        239 GGTATGGGACCATATTTTTAGAATTTACTTTGTAATTTCTATATCTTTCATTTTTTGAAA 298

Qy        148 TGATATTTACATTTTTATCAAAAAAAAAGT 177
              |||||||||||||| |||| ||||||| |||
Db        299 TGATATTTACATTCTTATAAAAAAAATAGT 328


RESULT 14
AQ964120/c
LOCUS       AQ964120                697 bp    DNA     linear   GSS 28-JAN-2000
DEFINITION  LERGT42TF LERG Arabidopsis thaliana genomic clone LERGT42, genomic
            survey sequence.
ACCESSION   AQ964120
VERSION     AQ964120.1  GI:6791709
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
```

```
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TF
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .697
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /strain="LANDSBERG ERECTA"
                      /db_xref="taxon:3702"
                      /clone="LERGT42"
                      /clone_lib="LERG"
                      /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                      sheared to 0.4-0.7 Kbp before ligation."
ORIGIN

  Query Match             3.3%;  Score 70.8;  DB 28;  Length 697;
  Best Local Similarity   86.7%;  Pred. No. 0.016;
  Matches   78;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy         88 GGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAAT 147
              ||  ||||  |||||  |||||||  ||||||||||||||||  ||  ||||||||||| |||||
Db        455 GGTATGGGACCATATTTTTAGAATTTACTTTGTAATTTCTATATCTTTCATTTTTTGAAA 396

Qy        148 TGATATTTACATTTTTATCAAAAAAAAAGT 177
              ||||||||||||| |||| |||||||| |||
Db        395 TGATATTTACATTCTTATAAAAAAAATAGT 366


RESULT 15
BX437758
LOCUS       BX437758               1200 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX437758 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YB01
            5-PRIME, mRNA sequence.
ACCESSION   BX437758
VERSION     BX437758.1  GI:30773605
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP008CA01QP1.
FEATURES              Location/Qualifiers
     source           1. .1200
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0CAP008YB01"
                      /tissue_type="THYMUS"
                      /clone_lib="Homo sapiens THYMUS"
                      /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                      with a NotI-oligo(dT) primer. Five prime end enriched,
                      double-strand cDNA was digested with Not I and cloned into
                      the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                      Library was not normalized."
ORIGIN

  Query Match             3.3%;  Score 70.6;  DB 13;  Length 1200;
  Best Local Similarity   32.6%;  Pred. No. 0.012;
  Matches  215;  Conservative  124;  Mismatches  318;  Indels    2;  Gaps    2;

Qy        618 AAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAAAGAACAATTTAATAAAG 677
              ||||  |  |||||  |  |  ||  |  ||||  |||       ||||  ||  ||  ||  |||
Db        207 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 266

Qy        678 TGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGGCCGCGACTCGTAACAAG 737
                  |  |||||||  :  :|   :   :|    | ::||   |  :   |||   |     |   :  :  ||:
Db        267 AAAAAAAAAAAAAWWAWAATWWWWAAAARATWDGATKADDATTGGGGGDWTTTWWWTWTAAR 326

Qy        738 GGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATTTTTTCTTTTTCTTTTCC 797
              :::  |       :     :|  |  :|  :  ||   |  :  |  :||||  :::::  :::
Db        327 RRDKTTTTTTTTTWAAAAARATAKRGGRGGADAWTAWWTWTWTTTTATWWWWWDDDDAKA 386

Qy        798 TAAAGAACGTTGTTTC-GTGCTTACAAGGGTCAAAACCATATCCAATTGTTCTGCCTATT 856
              :|:||||  ||  |||  :  |   :  ::|  :  :||:  ::  :  :|::  ::  :   | ||
Db        387 WADAGAATWTTTTTTWWRAGRGARKRDGTTDTRAADWWWWTWAAWAWWWDKRKTTTTTTT 446

Qy        857 ATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGTGATATATAATCTAACT 916
              :::  |  ||  |      |     |  |:  :          |:  |:  :  |  |  |::  :  :||
Db        447 TKKRRAAAAATTTTTTTTTTTTWTWAKAAAAAAARAATWTTWWRAAAAAAAWWTWWWAARG 506

Qy        917 TAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTGACCTCAGTCCCTAGTT 976
              |:  |  ||  |  ||:::  |  :|  :  |:|  |||  |  |  :  :  |   ||  ||
Db        507 RARGATTTTTTTTAAWWWGGGAGRTAWWTWTTWTTTTTTTTTTTTTTAWWWATWTATATTT 566

Qy        977 AGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAGTTCCTCTTCCTGCAAC 1036
                 :  :|:  |  |  |  ::|  :  |   :   |  :::  |  |  :  ||  |  |:       :
Db        567 TTWWTWAWAAAAASAGAWKATWTTTTKKTTTATWWKRAGGATWTTTTTTTTWTWAARGDW 626

Qy       1037 TAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTTGGCACTCTCGTATTAA 1096
              ::  |  ||  |  ||:  |  ||     |  |  ||:::  :  ::  :  |  :  :  |  |  ||  :|
Db        627 KWTWTWTTTTTTTWTTTTTTTTWTTGT-TTWWWRDGARRTAKATAKWWTTWTWWTAWAWA 685

Qy       1097 ATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATG 1156
                 |  :  :  |:       |  |::|  ||||   ||  :|  :|||   ||   |     ||  :|
Db        686 GGAARKRRATWTWTWAAAWAWWAGAAAARGAGAAGRAAWTTTTTTTATKAGARRGATWTT 745

Qy       1157 TTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTGAAGATATTTAT 1216
                :|::|  :       :  ||       ::|  ||  |  |  |:  |  ||::   ||  ::::||:|
Db        746 AWAWWTATRARGARGADTTTTTTATAWWTATTTTTTWWAWAAGATDKAAAAAWWWWTTWT 805

Qy       1217 ATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTGATGTAGATGT 1275
              |  |   ||  :  :  |  |  |:||:  |  :  |  |  |  |  :|:: ||  ||  |  ||  |
Db        806 AAAAAAAATTWWWAGARAAAKATWTWAAWAAAAAAAAAAAWRARWAATATATWTTTATAT 864

Search completed: July 31, 2004, 08:20:33
Job time : 5565 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 08:23:39 ; Search time 16500 Seconds
(without alignments)
11592.291 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatccacagtgaataaat.....tttcatgggttacgagtcca 4413

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		4413	100.0	4413	6	AX063411	AX063411 Sequence
2		4413	100.0	4413	8	AF242314	AF242314 Arabidops
3	c	4052	91.8	82454	8	AC012193	AC012193 Arabidops
4		2154	48.8	4309	6	AX063412	AX063412 Sequence
5		2149	48.7	2149	6	AX063409	AX063409 Sequence
6		2149	48.7	4280	6	AX063410	AX063410 Sequence
7		1486	33.7	1741	8	AF242313	AF242313 Arabidops
8	c	202	4.6	317	8	ATH553603	AJ553603 Arabidops
9		143	3.2	168	8	ATH522466	AJ522466 Arabidops
10		69	1.6	465	8	BT010186	BT010186 Arabidops
11		68	1.5	295	11	AL773135	AL773135 Arabidops
12	c	51	1.2	43982	8	AC007289	AC007289 Arabidops
13	c	26	0.6	15046	6	AX347002	AX347002 Sequence
14		26	0.6	124550	8	AC144806	AC144806 Medicago
15	c	26	0.6	125992	2	AC142222	AC142222 Medicago
16	c	26	0.6	137040	2	BX322621	BX322621 Danio rer
17	c	26	0.6	155937	2	AC016341	AC016341 Homo sapi
18	c	26	0.6	174741	2	AC024280	AC024280 Homo sapi
19		26	0.6	199200	9	AC010351	AC010351 Homo sapi
20		25	0.6	773	10	BC025131	BC025131 Mus muscu
21		25	0.6	1140	8	AF037457	AF037457 Fritillar
22		25	0.6	1777	8	AF179249	AF179249 Lycopersi
23		25	0.6	2033	6	BD190852	BD190852 Secreted
24	c	25	0.6	47573	3	AF030694	AF030694 Plasmodiu
25		25	0.6	125479	9	HSJ244F24	AL096865 Human DNA
26	c	25	0.6	145456	9	AL513366	AL513366 Human DNA
27		25	0.6	349751	3	PFMAL4P3	AL035476 Plasmodiu
28	c	24	0.5	646	11	BV065942	BV065942 S212P6179
29	c	24	0.5	7752	6	AX348656	AX348656 Sequence
30		24	0.5	8964	6	AX251977	AX251977 Sequence
31		24	0.5	8964	6	AX344369	AX344369 Sequence
32		24	0.5	8964	6	AX346437	AX346437 Sequence
33		24	0.5	8964	6	AX349014	AX349014 Sequence
34	c	24	0.5	87114	9	AC005353	AC005353 Homo sapi
35	c	24	0.5	91469	9	AL356963	AL356963 Human DNA
36	c	24	0.5	94605	9	AF410480S2	AF360695 Homo sapi
37		24	0.5	95851	9	HS1118M15	AL109964 Human DNA
38		24	0.5	109860	9	AC112695	AC112695 Homo sapi
39		24	0.5	110000	2	AC139250_4	Continuation (5 of
40	c	24	0.5	110000	2	PFMAL7P1_03	Continuation (4 of
41	c	24	0.5	115866	9	AL627422	AL627422 Human DNA
42		24	0.5	120994	9	AC008697	AC008697 Homo sapi
43		24	0.5	124949	9	AL358794	AL358794 Human DNA
44		24	0.5	125021	8	AC007504	AC007504 Arabidops
45	c	24	0.5	126467	9	AF487553	AF487553 Homo sapi

ALIGNMENTS

RESULT 1
AX063411
LOCUS AX063411 4413 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 3 from Patent WO0100833.
ACCESSION AX063411
VERSION AX063411.1 GI:12541199
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Hoffmann,B., Mollier,P. and Pelletier,G.
TITLE Promoter expressed specifically in the cells of plant roots,

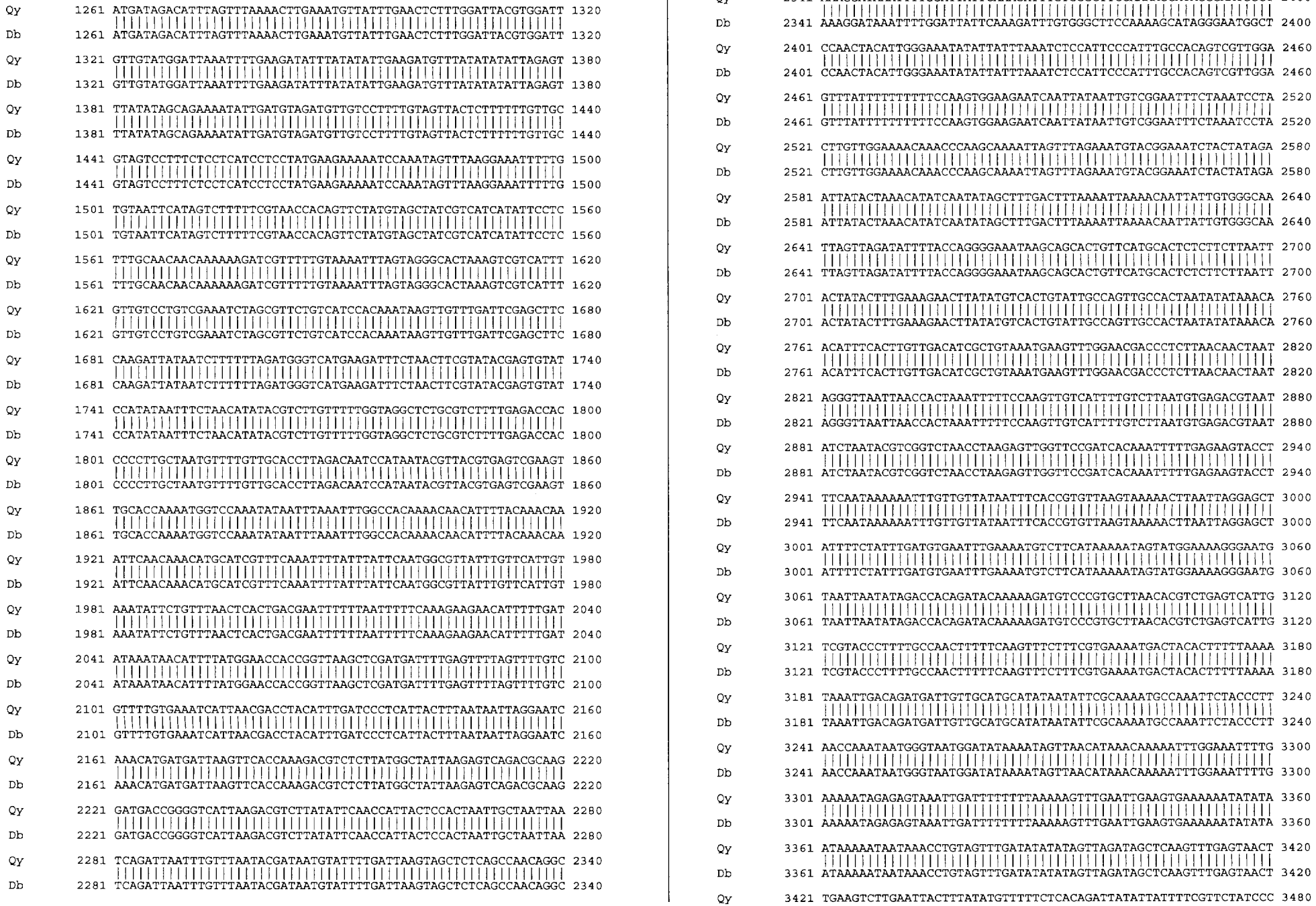
recombinant vectors and host cells comprising same and transgenic plants obtained
JOURNAL Patent: WO 0100833-A 3 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES Location/Qualifiers
source 1. .4413
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
ORIGIN

Query Match 100.0%; Score 4413; DB 6; Length 4413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60

Qy         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120

Qy        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180

Qy        181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240

Qy        241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300

Qy        301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360

Qy        361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420

Qy        421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480

Qy        481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540

Qy        541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600

Qy        601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660

Qy        661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720

Qy        721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780

Qy        781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 840

Qy        841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900

Qy        901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960

Qy        961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020

Qy       1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080

Qy       1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140

Qy       1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200

Qy       1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260

Qy       1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320

Qy       1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380

Qy       1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440

Qy       1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500

Qy       1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560

Qy       1561 TTTGCAACAACAAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTGCAACAACAAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620

Qy       1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680

Qy       1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740

Qy       1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800

Qy       1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860

Qy       1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920

Qy       1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980
```

```
Qy       1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040
Db       1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040

Qy       2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100
Db       2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100

Qy       2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160
Db       2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160

Qy       2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
Db       2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220

Qy       2221 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280
Db       2221 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280

Qy       2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340
Db       2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340

Qy       2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400
Db       2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400

Qy       2401 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 2460
Db       2401 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 2460

Qy       2461 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 2520
Db       2461 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 2520

Qy       2521 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 2580
Db       2521 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 2580

Qy       2581 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 2640
Db       2581 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 2640

Qy       2641 TTAGTTAGATATTTTACCAGGGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 2700
Db       2641 TTAGTTAGATATTTTACCAGGGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 2700

Qy       2701 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 2760
Db       2701 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 2760

Qy       2761 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 2820
Db       2761 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 2820

Qy       2821 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
Db       2821 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880

Qy       2881 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 2940
Db       2881 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 2940

Qy       2941 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 3000
Db       2941 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 3000

Qy       3001 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 3060
Db       3001 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 3060

Qy       3061 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 3120
Db       3061 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 3120

Qy       3121 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 3180
Db       3121 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 3180

Qy       3181 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 3240
Db       3181 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 3240

Qy       3241 AACCAAATAATGGGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTTG 3300
Db       3241 AACCAAATAATGGGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTTG 3300

Qy       3301 AAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATATA 3360
Db       3301 AAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATATA 3360

Qy       3361 ATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAACT 3420
Db       3361 ATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAACT 3420

Qy       3421 TGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATCCC 3480
Db       3421 TGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATCCC 3480

Qy       3481 AAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTTAAGAAGTCGGTGTATAAT 3540
Db       3481 AAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTTAAGAAGTCGGTGTATAAT 3540

Qy       3541 CTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTTTC 3600
Db       3541 CTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTTTC 3600

Qy       3601 AAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGGCG 3660
Db       3601 AAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGGCG 3660

Qy       3661 GGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCTCT 3720
Db       3661 GGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCTCT 3720

Qy       3721 GAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
Db       3721 GAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780

Qy       3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACATTT 3840
Db       3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACATTT 3840

Qy       3841 GGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGATTT 3900
Db       3841 GGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGATTT 3900

Qy       3901 CATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCGGG 3960
Db       3901 CATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCGGG 3960

Qy       3961 GTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
Db       3961 GTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020

Qy       4021 CTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGTCT 4080
Db       4021 CTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGTCT 4080

Qy       4081 ACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTTGT 4140
Db       4081 ACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTTGT 4140

Qy       4141 GGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGAAG 4200
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 GGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGAAG 4200

Qy       4201 CATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTTTT 4260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 CATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTTTT 4260

Qy       4261 AGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGTTT 4320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 AGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGTTT 4320

Qy       4321 GATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTATTT 4380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTATTT 4380

Qy       4381 TCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 4413
                ||||||||||||||||||||||||||||||||
Db       4381 TCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 4413


RESULT 2
AF242314
LOCUS       AF242314                4413 bp    DNA     linear   PLN 06-AUG-2000
DEFINITION  Arabidopsis thaliana ribosomal S13-like protein gene, partial cds;
            nuclear gene for chloroplast product.
ACCESSION   AF242314
VERSION     AF242314.1  GI:9719365
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4413)
  AUTHORS   Mollier,P., Hoffmann,B., Pelletier,G. and Chwetzoff,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Genetique, INRA, Route de St Cyr,
            Versailles 78026, France
FEATURES             Location/Qualifiers
     source          1. .4413
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="near m532"
                     /clone="BAC T32E8"
                     /ecotype="Columbia"
     mRNA            <3738. .>3806
                     /product="ribosomal S13-like protein"
     CDS             3738. .>3806
                     /note="similar to Arabidopsis thaliana ribosomal protein
                     S13 encoded by GenBank Accession Number Z47986"
                     /codon_start=1
                     /product="ribosomal S13-like protein"
                     /protein_id="AAF97767.1"
                     /db_xref="GI:9719366"
                     /translation="MLGLRRSATTLFDISQSLLRNVT"
ORIGIN

  Query Match             100.0%;  Score 4413;  DB 8;  Length 4413;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4413;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60

Qy         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120

Qy        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180

Qy        181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240

Qy        241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300

Qy        301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360

Qy        361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420

Qy        421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480

Qy        481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540

Qy        541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600

Qy        601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660

Qy        661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720

Qy        721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780

Qy        781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAAGATAAGAAGGTAGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAAGATAAGAAGGTAGGCA 840

Qy        841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900

Qy        901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960

Qy        961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020

Qy       1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080

Qy       1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140

Qy       1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200

Qy       1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260
```


Qy 1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320
Db 1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320
Qy 1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380
Db 1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380
Qy 1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440
Db 1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440
Qy 1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500
Db 1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500
Qy 1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560
Db 1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560
Qy 1561 TTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620
Db 1561 TTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620
Qy 1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680
Db 1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680
Qy 1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740
Db 1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740
Qy 1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800
Db 1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800
Qy 1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860
Db 1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860
Qy 1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920
Db 1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920
Qy 1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980
Db 1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980
Qy 1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040
Db 1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040
Qy 2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100
Db 2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100
Qy 2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160
Db 2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160
Qy 2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
Db 2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
Qy 2221 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280
Db 2221 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280
Qy 2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340
Db 2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340
Qy 2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400
Db 2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400
Qy 2401 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 2460
Db 2401 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 2460
Qy 2461 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 2520
Db 2461 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 2520
Qy 2521 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 2580
Db 2521 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 2580
Qy 2581 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 2640
Db 2581 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 2640
Qy 2641 TTAGTTAGATATTTTACCAGGGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 2700
Db 2641 TTAGTTAGATATTTTACCAGGGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 2700
Qy 2701 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 2760
Db 2701 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 2760
Qy 2761 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 2820
Db 2761 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 2820
Qy 2821 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
Db 2821 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
Qy 2881 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 2940
Db 2881 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 2940
Qy 2941 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 3000
Db 2941 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 3000
Qy 3001 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 3060
Db 3001 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 3060
Qy 3061 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 3120
Db 3061 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 3120
Qy 3121 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 3180
Db 3121 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 3180
Qy 3181 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 3240
Db 3181 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 3240
Qy 3241 AACCAAATAATGGGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTTG 3300
Db 3241 AACCAAATAATGGGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTTG 3300
Qy 3301 AAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATATA 3360
Db 3301 AAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATATA 3360
Qy 3361 ATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAACT 3420
Db 3361 ATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAACT 3420
Qy 3421 TGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATCCC 3480

```
Db       3421 TGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATCCC 3480

Qy       3481 AAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTAAGAAGTCGGTGTATAAT 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 AAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTAAGAAGTCGGTGTATAAT 3540

Qy       3541 CTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTTTC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTTTC 3600

Qy       3601 AAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGGCG 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 AAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGGCG 3660

Qy       3661 GGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCTCT 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCTCT 3720

Qy       3721 GAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 GAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780

Qy       3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACATTT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACATTT 3840

Qy       3841 GGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGATTT 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 GGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGATTT 3900

Qy       3901 CATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCGGG 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 CATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCGGG 3960

Qy       3961 GTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020

Qy       4021 CTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGTCT 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 CTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGTCT 4080

Qy       4081 ACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTTGT 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 ACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTTGT 4140

Qy       4141 GGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGAAG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 GGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGAAG 4200

Qy       4201 CATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTTTT 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 CATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTTTT 4260

Qy       4261 AGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGTTT 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 AGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGTTT 4320

Qy       4321 GATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTATTT 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTATTT 4380

Qy       4381 TCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 4413
             |||||||||||||||||||||||||||||||
Db       4381 TCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 4413


RESULT 3
AC012193/c
LOCUS       AC012193               82454 bp    DNA     linear   PLN 19-JAN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence,
            complete sequence.

ACCESSION   AC012193
VERSION     AC012193.6  GI:12323286
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 82454)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82454)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 82454)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280880.
            Address all correspondence to:at@tigr.org

            BAC clone T32E8 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .82454
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="T32E8"
     gene            complement(1. .3320)
                     /gene="T32E8.1"
                     /note="similar to ribonuclease II RNB family protein
                     GB:CAB16367 [Schizosaccharomyces pombe]; Pfam HMM hit:
                     VacB and RNase II family 3'-5' exoribonucleases"
     mRNA            complement(join(<1. .69,154. .2280,2381. .2449,2524. .2577,
                     2676. .3055,3215. .>3320))
                     /gene="T32E8.1"
     CDS             complement(join(<1. .69,154. .2280,2381. .2449,2524. .2577,
                     2676. .3055,3215. .3320))
                     /gene="T32E8.1"
                     /codon_start=1
                     /product="putative 3'-5' exoribonuclease, 3' partial;
                     3320-1"
                     /protein_id="AAG51632.1"
```

```
                     /db_xref="GI:12323300"
                     /translation="MKSASSEQSVERIENGHKKKRNRPQKQNRRSKQSSVPIEDAHVE
                     ESLDGRDSSRSKAKDSTSSSKQQRPNTDELEAMRASNVAFNSMPPMRAESGYPRRSAS
                     PLLSSPEVSKQLLSKSCPDPRACEQSPGMNGELFQQIEGSSQRKIFSSHWSLDAVTEA
                     LEKGEAFKALFRVNAHNRNEAYCKIDGVPTDILINGNVCQSRAVEGDTVVIKLDPLSL
                     WPKMKGFVTESAAKPEGTNSPPEKDDKKARQKNGIDVVEGFEDGFSKNKSSVIGKGAK
                     NGVTPSSPPSLDSCLGSFCEQKGNCSAVDKLCGILSSFPHKRPTGQVVAVVEKSLVRD
                     SIVGLLDVKGWIHYKESDPKRCKSPLSLSDDEYVQLMPADPRFPKLIVPFHVLPGSIR
                     ARLENLDPNLEAELVAAQIVDWGEGSPFPVAQITHLFGRGSELEPQINAILYQNSVCD
                     SDFSPGSLTSLPRVPWEVPEEEVQRRKDLRDLCVLTIDPSTATDLDDALSVQSLPGGF
                     FRVGVHIADVSYFVLPETALDTEARFRSTSVYLMQRKISMLPPLLSENVGSLSPGADR
                     LAFSILWDLNREGDVIDRWIGRTIIRSCCKLSYDHAQDIIDGKSDVAENGWPALHGSF
                     KWCDVTRSVKQLSEISTTLRQKRFRNGALQLENSKPVFLFDEHGVPYDFVTCSRKGSN
                     FLVEEFMLLANMTAAEVISQAYRASSLLRRHPEPNTRKLKEFEGFCSKHGMDLDISSS
                     GQLQDSLEKITGNLKDDSVFVDILNNYAIKPMQLASYFCTGNLKDSVAEWGHYALAVP
                     LYTHFTSPLRRYPDIVVHRALAAALEAEELYSKQKQTAIDEGRSCFTGIHFNKDAAES
                     IEGKEALSVAALKHGVPSTEILSDVAAYCNERKLAARKVRDACDKLYTWFVLKQKEIF
                     PCEARVMNLGSRFMTVYISKL"
misc_feature         1. .189
                     /note="overlap with BAC clone T5M16
                     [AC010704:103165. .103353]."
gene                 complement(8589. .10716)
                     /gene="T32E8.2"
                     /note="similar to AUX1 (regulator of root gravitropism,
                     putative permease) GB:CAA67308 [Arabidopsis thaliana]"
mRNA                 complement(join(<8589. .8783,9000. .9465,9555. .9652,
                     9753. .9936,10020. .10132,10236. .10421,10504. .>10716))
                     /gene="T32E8.2"
CDS                  complement(join(8589. .8783,9000. .9465,9555. .9652,
                     9753. .9936,10020. .10132,10236. .10421,10504. .10674))
                     /gene="T32E8.2"
                     /codon_start=1
                     /product="putative AUX1-like permease; 10674-8589"
                     /protein_id="AAG51630.1"
                     /db_xref="GI:12323298"
                     /translation="MAAEKIETVVAGNYLEMEREEENISGNKKSSTKTKLSNFFWHGG
                     SVYDAWFSCASNQVAQVLLTLPYSFSQLGMMSGILFQLFYGLMGSWTAYLISVLYVEY
                     RTRKEREKFDFRNHVIQWFEVLDGLLGKHWRNLGLIFNCTFLLFGSVIQLIACASNIY
                     YINDKLDKRTWTYIFGACCATTVFIPSFHNYRIWSFLGLAMTTYTSWYLTIASLLHGQ
                     AEDVKHSGPTTMVLYFTGATNILYTFGGHAVTVEIMHAMWKPQKFKAIYLLATIYVLT
                     LTLPSASAVYWAFGDKLLTHSNALSLLPKTGFRDTAVILMLIHQFITFGFASTPLYFV
                     WEKLIGVHETKSMFKRAMARLPVVVPIWFLAIIFPFFGPINSAVGSLLVSFTVYIIPA
                     LAHMLTFAPAPSRENAVERPPRVVGGWMGTYCINIFVVVWVFVVGFGFGGWASMVNFV
                     RQIDTFGLFTKCYQCPPHKP"
gene                 12104. .13574
                     /gene="T32E8.3"
                     /note="similar to thaumatin-like protein GB:AAB71214
                     [Arabidopsis thaliana]; Pfam HMM hit: Thaumatin family"
mRNA                 join(<12104. .12179,12505. .12628,12704. .>13574)
                     /gene="T32E8.3"
CDS                  join(12104. .12179,12505. .12628,12704. .13574)
                     /gene="T32E8.3"
                     /codon_start=1
                     /product="thaumatin-like protein; 12104-13574"
                     /protein_id="AAG51631.1"
                     /db_xref="GI:12323299"
                     /translation="MDRTIEIEFDINGPYDSLIRPNISVASLTSRSENNITFDPKIFS
                     SKDANSFRFVTKDLVTLSHHIIRVWITIRTPVNYSTGTKWSESARVFTIINSCDQTIW
                     PAITPGENFNGGGFELKPGQSIVFHAPVGWSGRIWGRTGCKFDSTGTGTCETGSCGST
                     LKCSASGKPPASLAEFTLAALDFYDVSLVDGFNLPMSVTPMNGKGNCSVAGCVADLRP
                     HCPQELAVKSNGKVISCRSACDVFDRDEYCCRGVYGNPVVCQPTYYSKIFKQACPTAY
                     SYAYDDPTSIMTCTASDYVISFCSSRFVLFTFLTTSHDYFNYDSNQVSFILDHDFYIY
                     MIYFIHPLLISYAKVKVIQMQK"
gene                 14054. .15303
                     /gene="T32E8.4"
                     /note="similar to hypothetical protein GB:P34661
                     [Caenorhabditis elegans]"
mRNA                 join(14054. .14174,14251. .14348,14779. .14811,15170. .15303)
                     /gene="T32E8.4"
CDS                  join(14107. .14174,14251. .14348,14779. .14811,15170. .15252)
                     /gene="T32E8.4"
                     /codon_start=1
                     /product="unknown protein; 14107-15252"
                     /protein_id="AAG51633.1"
                     /db_xref="GI:12323301"
                     /translation="MATGGKVSFKVTLTSDPKLPFKVFSVPEGAPFTAVLKFAAEEFK
                     VPPQTSAIITNDGIGINPQQSAGNVFLKHGSELRLIPRDRVGAVFVMDP"
gene                 18583. .21388
                     /gene="T32E8.5"
                     /note="Pfam HMM hit: Eukaryotic protein kinase domain,
                     C-terminal region is identical to putative protein kinase
                     GB:CAA72680 [Arabidopsis thaliana]"
mRNA                 join(<18583. .18916,19056. .19948,20068. .20268,20382. .20543,
                     20662. .21388)
                     /gene="T32E8.5"
CDS                  join(18583. .18916,19056. .19948,20068. .20268,20382. .20543,
                     20662. .21234)
                     /gene="T32E8.5"
                     /codon_start=1
                     /product="putative protein kinase; 18583-21234"
                     /protein_id="AAG51619.1"
                     /db_xref="GI:12323287"
                     /translation="MQTTSIGPRRSVAPQRQASRNTRLVTAEGQRSQDVVTLSQSLAA
                     NTLTQDTTNNLAITSVAGESASITQPTVSEHFNPSDRQMDFGKSAVTSLESNLDVQRK
                     SQSLIGTSQDMEWDATNQAEASHLDACIGSKHQNLPSVDSEVSLKSEYKDSSSLAKIQ
                     GQLGEFPNFLNQPRTRCSAVGSSWATTTLIHSSSAPMLNATTHVSRSYVEADSNANPH
                     AVQSQGNLPSCCPSSKVSNILHPNKDATASEMPASTNDPEVRVKETDTSKQQQITTGL
                     EAPVGSSIYGSDGQANARLPEELHTSVSSQPQKSDKHEKVASSKGPSAPRKRNYDPDL
                     FFKVNGKLYQRLGKIGSGGSSEVHKVISSDCTIYALKKIKLKGRDYATAYGFCQEIGY
                     LKKLKGKTNIIQLIDYEVTDKTLLQEVLNGTMSNKDGRVKEDGFIYMVLEYGEIDLAH
                     MLSQKWREIEGSDRTIDENWLRFYWQQILQAVNTIHEERIVHSDLKPANFLLVRGFLK
                     LIDFGIAKAINSDTTNIQRDSQVGTLSYMSPEAFMCNESDENGNTIKCGRPSDIWSLG
                     CILYQMVYGRTPFADYKTFWAKFKVITDPNHEITYNQLSNPWLIDLMKKCLAWDRNQR
                     WRIPELLQHPFLAPPIPHEPQVKTIKLFSLIAESCGSDDDKANSMISQLEQLLSNPAP
                     LPRNDVLDSRDQNQQLLSRVSELCIQLQDRLQGLEGEN"
gene                 21802. .22921
                     /gene="T32E8.6"
                     /note="similar to putative SWH1 protein GB:AAC20736
                     [Arabidopsis thaliana]; Pfam HMM hit: PH (pleckstrin
                     homology) domain"
mRNA                 join(<21802. .22203,22526. .>22921)
                     /gene="T32E8.6"
CDS                  join(21802. .22203,22526. .22921)
                     /gene="T32E8.6"
                     /codon_start=1
                     /product="hypothetical protein; 21802-22921"
                     /protein_id="AAG51621.1"
                     /db_xref="GI:12323289"
                     /translation="MSSPLAAKSNLTTTRSASLNIPHSTRTNHNHDLRYSLSAGPRTN
                     EDRPASGNGVAGILYKWVNYGQGWKRRWFVLQDGVLSYYRIHGPDKISLSVEMDRRSK
                     LIGGESLRFICRHSKRGDVHSPGKPLGQIHLKVSSIGQSISDGKRFTVFTGTKSLHLR
                     AATSEDRASWIEALKAVKETFPRMSNEELMASTTNVSVSTDKLRQRLMEEEVDETIIK
                     DCEDIMKNNFLALHDEVMSLKRYQYHLIDSLKNVSNSPIRPSNQSFS"
gene                 27895. .31218

  Query Match           91.8%;  Score 4052;  DB 8;  Length 82454;
  Best Local Similarity 99.9%;  Pred. No. 0;
  Matches 4412;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42585 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 42526

Qy         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42525 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 42466

Qy        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42465 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 42406

Qy        181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42405 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 42346

Qy        241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300
```

```
Db      42345 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 42286

Qy        301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360

Db      42285 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 42226

Qy        361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420

Db      42225 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 42166

Qy        421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480

Db      42165 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 42106

Qy        481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540

Db      42105 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 42046

Qy        541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600

Db      42045 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 41986

Qy        601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660

Db      41985 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 41926

Qy        661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720

Db      41925 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 41866

Qy        721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780

Db      41865 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 41806

Qy        781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 840

Db      41805 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 41746

Qy        841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900

Db      41745 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 41686

Qy        901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960

Db      41685 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 41626

Qy        961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020

Db      41625 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 41566

Qy       1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080

Db      41565 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 41506

Qy       1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140

Db      41505 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 41446

Qy       1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200

Db      41445 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 41386

Qy       1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260

Db      41385 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 41326

Qy       1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320

Db      41325 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 41266

Qy       1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380

Db      41265 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 41206

Qy       1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440

Db      41205 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 41146

Qy       1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500

Db      41145 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 41086

Qy       1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560

Db      41085 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 41026

Qy       1561 TTTGCAACAACAAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620

Db      41025 TTTGCAACAACAAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 40966

Qy       1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680

Db      40965 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 40906

Qy       1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740

Db      40905 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 40846

Qy       1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800

Db      40845 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 40786

Qy       1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860

Db      40785 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 40726

Qy       1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920

Db      40725 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 40666

Qy       1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980

Db      40665 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 40606

Qy       1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040

Db      40605 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 40546

Qy       2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100

Db      40545 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 40486

Qy       2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160

Db      40485 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 40426

Qy       2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220

Db      40425 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 40366

Qy       2221 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280

Db      40365 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 40306

Qy       2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340

Db      40305 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 40246

Qy       2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400

Db      40245 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 40186

Qy       2401 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 2460

Db      40185 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 40126
```

```
Qy       2461 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40125 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 40066

Qy       2521 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40065 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 40006

Qy       2581 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40005 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 39946

Qy       2641 TTAGTTAGATATTTTACCAGGGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 2700
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      39945 TTAGTTAGATATTTTACCA-GGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 39887

Qy       2701 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39886 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 39827

Qy       2761 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39826 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 39767

Qy       2821 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39766 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 39707

Qy       2881 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39706 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 39647

Qy       2941 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39646 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 39587

Qy       3001 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39586 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 39527

Qy       3061 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39526 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 39467

Qy       3121 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39466 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 39407

Qy       3181 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39406 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 39347

Qy       3241 AACCAAATAATG-GGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTT 3299
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      39346 AACCAAATAATGTGGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTT 39287

Qy       3300 GAAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATAT 3359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39286 GAAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATAT 39227

Qy       3360 -AATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAA 3418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39226 AAATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAA 39167

Qy       3419 CTTGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATC 3478
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39166 CTTGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATC 39107

Qy       3479 CCAAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTAAGAAGTCGGTGTATA 3538
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39106 CCAAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTAAGAAGTCGGTGTATA 39047

Qy       3539 ATCTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTT 3598
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39046 ATCTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTT 38987

Qy       3599 TCAAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGG 3658
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38986 TCAAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGG 38927

Qy       3659 CGGGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCT 3718
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38926 CGGGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCT 38867

Qy       3719 CTGAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACAT 3778
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38866 CTGAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACAT 38807

Qy       3779 CAGCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACAT 3838
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38806 CAGCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACAT 38747

Qy       3839 TTGGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGAT 3898
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38746 TTGGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGAT 38687

Qy       3899 TTCATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCG 3958
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38686 TTCATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCG 38627

Qy       3959 GGGTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAG 4018
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38626 GGGTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAG 38567

Qy       4019 TGCTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGT 4078
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38566 TGCTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGT 38507

Qy       4079 CTACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTT 4138
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38506 CTACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTT 38447

Qy       4139 GTGGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGA 4198
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38446 GTGGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGA 38387

Qy       4199 AGCATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTT 4258
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38386 AGCATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTT 38327

Qy       4259 TTAGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGT 4318
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38326 TTAGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGT 38267

Qy       4319 TTGATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTAT 4378
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38266 TTGATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTAT 38207

Qy       4379 TTTCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 4413
             |||||||||||||||||||||||||||||||||||
Db      38206 TTTCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 38172
```

RESULT 4
AX063412

```
LOCUS       AX063412               4309 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 4 from Patent WO0100833.
ACCESSION   AX063412
VERSION     AX063412.1  GI:12541200
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Hoffmann,B., Mollier,P. and Pelletier,G.
```

TITLE Promoter expressed specifically in the cells of plant roots, recombinant vectors and host cells comprising same and transgenic plants obtained
JOURNAL Patent: WO 0100833-A 4 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)

FEATURES Location/Qualifiers
source 1. .4309
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Insert du vecteur pBin19"

ORIGIN

Query Match 48.8%; Score 2154; DB 6; Length 4309;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   131 CTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAAC 190
Db    25 CTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAAC 84

Qy   191 AACCGGCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTAC 250
Db    85 AACCGGCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTAC 144

Qy   251 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 310
Db   145 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 204

Qy   311 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 370
Db   205 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 264

Qy   371 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 430
Db   265 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 324

Qy   431 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 490
Db   325 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 384

Qy   491 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 550
Db   385 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 444

Qy   551 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 610
Db   445 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 504

Qy   611 TTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 670
Db   505 TTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 564

Qy   671 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 730
Db   565 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 624

Qy   731 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 790
Db   625 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 684

Qy   791 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACG 850
Db   685 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACG 744

Qy   851 TATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 910
Db   745 TATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 804

Qy   911 TGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCA 970
Db   805 TGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCA 864

Qy   971 TATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTA 1030
Db   865 TATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTA 924

Qy  1031 TTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTC 1090
Db   925 TTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTC 984

Qy  1091 TACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTG 1150
Db   985 TACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTG 1044

Qy  1151 AAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGA 1210
Db  1045 AAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGA 1104

Qy  1211 ATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACA 1270
Db  1105 ATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACA 1164

Qy  1271 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1330
Db  1165 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1224

Qy  1331 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1390
Db  1225 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1284

Qy  1391 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTT 1450
Db  1285 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTT 1344

Qy  1451 CTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCAT 1510
Db  1345 CTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCAT 1404

Qy  1511 AGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAA 1570
Db  1405 AGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAA 1464

Qy  1571 CAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGT 1630
Db  1465 CAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGT 1524

Qy  1631 CGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATA 1690
Db  1525 CGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATA 1584

Qy  1691 ATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1750
Db  1585 ATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1644

Qy  1751 TCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTA 1810
Db  1645 TCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTA 1704

Qy  1811 ATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAA 1870
Db  1705 ATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAA 1764

Qy  1871 TGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAA 1930
Db  1765 TGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAA 1824

Qy  1931 CATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTG 1990
Db  1825 CATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTG 1884

Qy  1991 TTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACA 2050
Db  1885 TTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACA 1944

Qy  2051 TTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAA 2110
```

```
Db       1945 TTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAA 2004

Qy       2111 ATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATG 2170

Db       2005 ATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATG 2064

Qy       2171 ATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2230

Db       2065 ATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2124

Qy       2231 GTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284

Db       2125 GTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2178


RESULT 5
AX063409
LOCUS       AX063409               2149 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 1 from Patent WO0100833.
ACCESSION   AX063409
VERSION     AX063409.1  GI:12541197
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Hoffmann,B., Mollier,P. and Pelletier,G.
  TITLE     Promoter expressed specifically in the cells of plant roots,
            recombinant vectors and host cells comprising same and transgenic
            plants obtained
  JOURNAL   Patent: WO 0100833-A 1 04-JAN-2001;
            INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES             Location/Qualifiers
     source          1. .2149
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

  Query Match             48.7%;  Score 2149;  DB 6;  Length 2149;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195

Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy        196 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255

Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315

Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375

Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435

Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495

Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555

Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy        556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615

Db        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy        616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675

Db        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy        676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735

Db        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy        736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795

Db        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy        796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855

Db        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy        856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915

Db        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy        916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975

Db        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy        976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035

Db        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy       1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095

Db        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy       1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155

Db        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy       1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215

Db       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy       1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275

Db       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy       1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335

Db       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy       1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395

Db       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy       1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455

Db       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy       1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515

Db       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy       1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575

Db       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy       1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635

Db       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
```

```
Qy       1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy       1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy       1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy       1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy       1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy       1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

```
RESULT 6
AX063410
LOCUS       AX063410                4280 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 2 from Patent WO0100833.
ACCESSION   AX063410
VERSION     AX063410.1  GI:12541198
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Hoffmann,B., Mollier,P. and Pelletier,G.
  TITLE     Promoter expressed specifically in the cells of plant roots,
            recombinant vectors and host cells comprising same and transgenic
            plants obtained
  JOURNAL   Patent: WO 0100833-A 2 04-JAN-2001;
            INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES             Location/Qualifiers
     source          1. .4280
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Construction promoteur + sequence codante du g ne
                     gus"
ORIGIN

  Query Match             48.7%;  Score 2149;  DB 6;  Length 4280;
  Best Local Similarity   100.0%;  Pred. No. 0;
```

```
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy        196 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy        556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy        616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy        676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy        736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy        796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy        856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy        916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy        976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy       1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy       1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy       1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
```

```
Qy       1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy       1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy       1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy       1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy       1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy       1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy       1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy       1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy       1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy       1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy       1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy       1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy       1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

```
RESULT 7
AF242313
LOCUS       AF242313                1741 bp    DNA     linear   PLN 06-AUG-2000
DEFINITION  Arabidopsis thaliana unknown sequence.
ACCESSION   AF242313
VERSION     AF242313.1  GI:9719364
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1741)
  AUTHORS   Mollier,P., Hoffmann,B., Pelletier,G. and Chwetzoff,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Genetique, INRA, Route de St Cyr,
            Versailles 78026, France
FEATURES             Location/Qualifiers
     source          1. .1741
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="NIA1-agp64"
                     /ecotype="Wassilewskija"
     misc_feature    <1. .>1741
                     /note="contains root-specific promoter"
ORIGIN

  Query Match             33.7%;  Score 1486;  DB 8;  Length 1741;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1736;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        544 AATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGG 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGG 60

Qy        604 TTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAAT 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAAT 120

Qy        664 TTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAG 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAG 180

Qy        724 GGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGG 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGG 240

Qy        784 AAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAA 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAA 300

Qy        844 GAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGA 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGA 360

Qy        904 CGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTC 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTC 420

Qy        964 AAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTT 1023
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAACCATATCCAATTTTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTT 480

Qy       1024 GTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCT 1083
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCT 540

Qy       1084 ATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCA 1143
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCA 600
```

```
Qy       1144 AAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAA 1203
             ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATTTGATTAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAA 660

Qy       1204 ATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATG 1263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATG 720

Qy       1264 ATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTT 1323
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTT 780

Qy       1324 GTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTA 1383
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTA 840

Qy       1384 TATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTA 1443
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTA 900

Qy       1444 GTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGT 1503
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGT 960

Qy       1504 AATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTT 1563
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTT 1020

Qy       1564 GCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTT 1623
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GCCACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTT 1080

Qy       1624 GTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAA 1683
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAA 1140

Qy       1684 GATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCA 1743
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCA 1200

Qy       1744 TATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCC 1803
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCC 1260

Qy       1804 CTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGC 1863
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGC 1320

Qy       1864 ACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATT 1923
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATT 1380

Qy       1924 CAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAA 1983
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAA 1440

Qy       1984 TATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATA 2043
             ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1441 TATTCTGTTTAACTCACTGACGAATTTTTTATTTTTTCAAAGAAGAACATTTTTGATATA 1500

Qy       2044 AATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTT 2103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTT 1560

Qy       2104 TTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAA 2163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAA 1620

Qy       2164 CATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGAT 2223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGAT 1680
```

```
Qy       2224 GACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCA 2283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCA 1740

Qy       2284 G 2284
             |
Db       1741 G 1741
```

```
RESULT 8
ATH553603/c
LOCUS       ATH553603               317 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            362F10.
ACCESSION   AJ553603
VERSION     AJ553603.1  GI:29369764
KEYWORDS    left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
  MEDLINE   22363535
   PUBMED   12446565
REFERENCE   2  (bases 1 to 317)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Wassillewskija"
                     /db_xref="taxon:3702"
                     /clone="362F10"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
     misc_feature    1. .317
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN

  Query Match             4.6%;  Score 202;  DB 8;  Length 317;
  Best Local Similarity   99.6%;  Pred. No. 6.2e-91;
  Matches  252;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1888 AAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGCATCGTTTCAAAT 1947
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 AAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGCATCGTTTCAAAT 258

Qy       1948 TTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAACTCACTGACGAA 2007
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 TTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAACTCACTGACGAA 198

Qy       2008 TTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACCACCGG 2067
```

```
                ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 TTTTTTATTTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACCACCGG 138

Qy       2068 TTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCATTAACGACCTACA 2127
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 TTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCATTAACGACCTACA 78

Qy       2128 TTTGATCCCTCAT 2140
                ||||||||||||
Db         77 TTTGATCCCTCAT 65


RESULT 9
ATH522466
LOCUS       ATH522466               168 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            292E05.
ACCESSION   AJ522466
VERSION     AJ522466.1  GI:26790702
KEYWORDS    left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
  MEDLINE   22363535
   PUBMED   12446565
REFERENCE   2  (bases 1 to 168)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .168
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Wassillewskija"
                     /db_xref="taxon:3702"
                     /clone="292E05"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
     misc_feature    1. .168
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN

  Query Match             3.2%;  Score 143;  DB 8;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-61;
  Matches  143;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2472 TTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTACTTGTTGGAAA 2531
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 TTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTACTTGTTGGAAA 85

Qy       2532 ACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGAATTATACTAAA 2591
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 ACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGAATTATACTAAA 145

Qy       2592 CATATCAATATAGCTTTGACTTT 2614
                |||||||||||||||||||||||
Db        146 CATATCAATATAGCTTTGACTTT 168


RESULT 10
BT010186
LOCUS       BT010186                465 bp    mRNA    linear   PLN 12-AUG-2003
DEFINITION  Arabidopsis thaliana At1g77750 gene, complete cds.
ACCESSION   BT010186
VERSION     BT010186.1  GI:33589777
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
            Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 465)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
            Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-2003) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Cheuk,R.,
            Chen,H., Kim,C.J., Shinn,P., Bowser,L., Chan, M.M., Chang,C.M.,
            Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera, C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
            Yamada,K., Yu,G., Yuan,S., Davis,R.W.,  Theologis,A., and
            Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="U61319"
                     /ecotype="Columbia"
                     /note="This clone is in pUNI 51"
     CDS             1. .465
                     /note="unknown protein"
                     /codon_start=1
                     /product="At1g77750"
```

```
                     /protein_id="AAQ22655.1"
                     /db_xref="GI:33589778"
                     /translation="MLGLRRSATTLFDISQSLLRNVTFHGLRVQGIRVGNAEVPNNKP
                     LKTGLQEVYGIGRRKSHQVLCHLGITNKLARDLTGKELIDLREEVGQHQHGDELRRRV
                     GSEIQRLVEVDCYRGSRHRHGLPCRGQRTSTNARTKKGKAVAIAGKKKAPRK"
ORIGIN

  Query Match             1.6%;  Score 69;  DB 8;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-23;
  Matches   69;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3738 ATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCAGCCAGTCTCTGCTTCGT 3797
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db             1 ATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCAGCCAGTCTCTGCTTCGT 60

Qy       3798 AATGTTACG 3806
             |||||||||
Db            61 AATGTTACG 69


RESULT 11
AL773135
LOCUS       AL773135                295 bp    DNA     linear   STS 11-JUN-2003
DEFINITION  Arabidopsis thaliana transposon insertion STS SM_3.1249, sequence
            tagged site.
ACCESSION   AL773135
VERSION     AL773135.1  GI:21535322
KEYWORDS    STS; STS, sequence tagged site.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B.,
            Murphy,G., Langham,S., LeGrys,C., Jones,J.D.G. and Bevan,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 295)
  AUTHORS   Clarke,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-2002) Clarke J.H., John Innes Centre, Colney
            Lane, Norwich, NR4 7UJ, UK
COMMENT     AT denotes an activation tag dissociation transposon within a
            single line, ET an enhancer trap dissociation transposon, GT a gene
            trap dissociation transposon, MT a mis-expression enhancer trap
            dissociation transposon, SM a defective suppressor mutator
            transposon. _3 denotes a sequence derived from the 3'end of the
            transposon, _5 denotes a sequence derived from the 5'end of the
            transposon BBSRC GARNet, ATIS project
            On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
            code: N55909.
FEATURES             Location/Qualifiers
     source          1. .295
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia-0 NASC stock code N1092"
                     /db_xref="taxon:3702"
                     /clone="AC012193"
                     /note="Derived from superpool 12.77 NASC code Unknown"
     STS             1. .295
                     /standard_name="SM_3.1249"
ORIGIN

  Query Match             1.5%;  Score 68;  DB 11;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-22;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        456 ACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAA 515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAAGTGGTATATTGTGAAAAAAA 60

Qy        516 AAAAAAAA 523
              ||||||||
Db         61 AAAAAAAA 68


RESULT 12
AC007289/c
LOCUS       AC007289               43982 bp    DNA     linear   PLN 27-FEB-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone F16J10 map RNS1, complete
            sequence.
ACCESSION   AC007289
VERSION     AC007289.9  GI:20198243
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 43982)
  AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
            Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
            Nierman,W.C., Fraser,C.M. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43982)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 43982)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598749.
            Information on performance of analysis and a more detailed
             annotation of this entry and other sequences of chromosomes 3, 4
             and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .43982
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="RNS1"
                     /clone="F16J10"
     misc_feature    1. .5724
                     /note="overlap with BAC clone T20G20
                     (AC006220:104629. .110352)."
     misc_feature    1
                     /note="17532 nt  before this point were not included in
                     the submitted sequence, due to overlap with another BAC"
     repeat_region   complement(1434. .1472)
                     /rpt_family="AT_rich"
     repeat_region   complement(1907. .1945)
                     /rpt_family="AT_rich"
     gene            complement(2734. .3407)
                     /gene="At2g05435"
                     /note="synonym: F16J10.1; pseudogene"
                     /pseudo
     repeat_region   complement(3455. .3489)
                     /rpt_family="(TA)n"
     repeat_region   complement(3808. .3836)
                     /rpt_family="AT_rich"
     repeat_region   complement(3859. .3914)
                     /rpt_family="Rf:ATREP9|ATREP9 ATREP9 repeat - a
                     consensus."
     repeat_region   complement(3978. .4108)
                     /rpt_family="AT_rich"
     repeat_region   complement(4754. .4787)
                     /rpt_family="AT_rich"
     gene            complement(5500. .6147)
                     /gene="At2g05430"
                     /note="synonym: F16J10.2; predicted by genefinder"
```

```
mRNA            complement(join(<5500. .5949,6031. .>6147))
                /gene="At2g05430"
CDS             complement(join(5500. .5949,6031. .6147))
                /gene="At2g05430"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAM15480.1"
                /db_xref="GI:20198245"
                /translation="MSKISESKSTLSHHGTAYDEDLDNDEIVNIKGFWIFRSKLGLAK
                RVFEKHPETTTKFCLKSEFAKETYLTALLDLIDMINMLPQQSLFEAELKEAENTILDL
                EAAGFKLDWLKRKLEEIRVTKKKAKNRTTRMRELDRKIQKHLEELSVLQEEMKKEQFE
                AMCDKPEYYLDSLCLNEVTCDSSEPITK"
gene            complement(7260. .8700)
                /gene="At2g05420"
                /note="synonym: F16J10.3; predicted by genefinder"
mRNA            complement(join(<7260. .7640,7857. .8077,8329. .8510,
                8591. .>8700))
                /gene="At2g05420"
CDS             complement(join(7260. .7640,7857. .8077,8329. .8510,
                8591. .8700))
                /gene="At2g05420"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAD32895.1"
                /db_xref="GI:4914358"
                /translation="MGKQVDSKTITWVIENFSSLQSASIHSDQFVVGDCKWRLKAYPK
                GNEKATYLAYRANNLALYLNVANSKSFPIGWTRHTKFSLTLVNQKSEKLSKLTESQHW
                FDHKSTSRGFPAMIPLTNLHTNEGFLVNGELTLVAKVEVLEVVGKLDVSKKSSPVMET
                IDVNGFQVLPQIESVNRLFAKHLDIASKFRPKKPYMKTAYMNVLLSLTKTLCQSPQDL
                SNDDISGAGAALTYLREAGFKLDWLEKKHGEIKEKKKKEEASLKRLQEMEKQIFNEAQ
                IYKEKVLAARAPLSLNEDNVF"
repeat_region   9262. .9302
                /rpt_family="AT_rich"
repeat_region   complement(9910. .9970)
                /rpt_family="AT_rich"
repeat_region   complement(10000. .10019)
                /rpt_family="(TAAA)n"
repeat_region   10066. .10101
                /rpt_family="AT_rich"
gene            complement(14048. .15111)
                /gene="At2g05410"
                /note="synonym: F16J10.4; predicted by genscan"
mRNA            complement(join(<14048. .14464,14624. .14847,14955. .>15111))
                /gene="At2g05410"
CDS             complement(join(14048. .14464,14624. .14847,14955. .15111))
                /gene="At2g05410"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAD32896.1"
                /db_xref="GI:4914359"
                /translation="MAYPKGINKAHDSFSLFLNVPDNESLPTGWRRHAKVSFSLVNQG
                SEKLSQRKVTQHWFVQKAFTWGFPVMITHTELNAKMGFLVNGELKVVAKIEVLEVVGK
                LDVSKESSPIMKTIDVNGFQVLPSQVDSVKRLFEKNLDIVSKFRLKNPYLKTACMNLL
                LSLTETLCQSPQELSNDGLSDAGVALAYLIETGLKLDWLEKKLDELKEKKKKEESCLV
                RLREMDEQLQPFKKRCLDIEDQISKEKEELLAAREPLSLYDDIDNNV"
repeat_region   complement(15873. .15972)
                /rpt_family="AT_rich"
gene            complement(16262. .17096)
                /gene="At2g05400"
                /note="synonym: F16J10.5; predicted by genscan"
mRNA            complement(join(<16262. .16672,16854. .>17096))
                /gene="At2g05400"
CDS             complement(join(16262. .16672,16854. .17096))
                /gene="At2g05400"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAD32897.1"
                /db_xref="GI:4914360"
                /translation="MFCDSAETQQWFDKNAPDWGFTEMISLTELHAEEGLLRNGELTV
                VAKVEVLEVVGKLDVSEESSPVIETIDVNGFQVLPSQVEYAKSLFERHLDIASKFRPK
                NPYLKTAYMNVLLSLTQTICQSPQELSNDDLSDAGAALAYLREAGFELDWLEKKLNEV
                KEKKKKEEACLAEIQDMDEHVKPLKKKYLDLEAQIDKKKAELLAARAPLSLNDDNVV"
gene            complement(17875. .21892)
                /gene="At2g05390"
                /note="synonym: F16J10.6"
mRNA            complement(join(<17875. .19791,19831. .20615,20671. .>21892))
                /gene="At2g05390"
CDS             complement(join(17875. .19791,19831. .20615,20671. .21892))
                /gene="At2g05390"
                /codon_start=1
                /product="putative retroelement pol polyprotein"
                /protein_id="AAD32898.1"
                /db_xref="GI:4914361"
                /translation="MEATLRVHKVWETIDPGSDDMEKNDMARALLFQSVPESTILQVG
                KHKTSKAMWEAIKTRNLGAERVKEAKLQTLMAEFDRLNMKDNETIDEFVGRISEISTK
                SESLGEEIEESKIVKKFLKSLPRKKYIHIIAALEQILDLNTTGFEDIVGRMKTYEDRV
                CDEDDSPEEQGKLMYANSESSYDTRGGRGRGRGRSSGRGRGGYGYQQRDKSKVICYRC
                DKTGHYASECLDRLLKLIKAQEQQQNNEDDDEIESLMMHEVVYLNERSVKPKEFEACS
                DNSWYLDNGASNHMTGNLQWFSKLNEMITGKVRFGDDSRIDIKGKGSIVLITKGGIRK
                TLTDVYFIPDLKSNIISLGQATEAGCDVRMKDDQLTLHDREGCLLLRATRSRNRLYKV
                DLNVENVKCLQLEAATMVRKELVIGISNIPKEKETCGSCLLGKQARQPFPKATTYRAS
                QVLELVHGDLCGPITQSTTAKKRYILVLIDDHTRYMWSMLLKEKSEAFEKFRDFKTKV
                EQESGVKIKTFRTDKGGEFVSQEFQDFCAKEGINRHLTAPYTPQQNGVVERRNRTLLG
                MTRSILKHMKMPNYLWGEAVRHSTYIINRVGTRSLQNQTPYEVFKQRKPNVEHLRVFG
                CIGYAKIEGPHLRKLDDRSKMLVYLGTEPGSKAYRLLDPTNRKIIKWNNSDSETRDIS
                GTFSLTLGEFGNNGIQESDDIETEKNGEESENSHEEEGENEHNEQEQIDAEETQPSHA
                TPLPTLRRSTRQVGKPNYLDDYVLMAEIEGEQVLLAINDEPWDFKEANKLKEWRDACK
                EEILSIEKNKTWSLIDLPVRRKVIGLKWVFKIKRNSDGSINKYKARLVAKGYVQRHGI
                DYDEVFAHVARIETIRVIIALAASNGWEVHHLDVKTAFLHGELREDVYVTQPEGFTNK
                DNEGKVYKLHKALYGLKQAPRAWNTKLNKILQELNFVKCSKEPSVYRRQEEKKLLIVA
                IYVDDLLVTGSSLDLILCFKKDMAGKFEMSDLGQLTYYLGIEVLHRKNGIILRQERYA
                MKIIEEAGMSNCNPVLIPMAAGLELCKAQEEKCITERDYRRMIGCLRYIVHTRPDLSY
                CVGVLSRYLQQPRESHGNALKQVLRYLKGTMSHGLYLKRGFKSGLVGYSDSSHSADLD
                DGKSTAGHIFYLHQCPITWCSQKQQVVALSSCEAEFMAATEAAKQAIWLQDLFAEVCG
                TTSEKVMIRVDNKSAIALTKNLVFHGRSKHIHRRYHFIRECVENNLVEVDHVPGVEQR
                ADILTKPLGRIKFREMRELVGVQCVLEDDFKLKGEFVGISMKSIN"
repeat_region   complement(22479. .22502)
                /rpt_family="AT_rich"
gene            complement(24570. .25781)
                /gene="At2g05380"
                /note="synonym: F16J10.7; supported by cDNA:
                gi_15146251_gb_AY049267.1_"
mRNA            complement(join(<24570. .25058,25678. .>25781))
                /gene="At2g05380"
CDS             complement(join(24781. .25058,25678. .25750))
                /gene="At2g05380"
                /codon_start=1
                /product="expressed protein"
                /protein_id="AAD32899.2"
                /db_xref="GI:20198244"
                /translation="MASKTLLLLGLFAFLFIVSEMAAAGTVKSESEETVKPEQHGGGF
                GDNGGGRYQGGGGHGGHGGGGYQGGGGRYQGGGGRQGGGGSYCRHGCCYKGYHGCSRC
                CSYAGEAVQTQSGH"
repeat_region   complement(24878. .24990)
                /rpt_family="(GGA)n"
repeat_region   complement(25118. .25153)
                /rpt_family="AT_rich"
repeat_region   complement(26299. .26329)
                /rpt_family="(A)n"
repeat_region   complement(28917. .28939)
                /rpt_family="AT_rich"
repeat_region   28929. .29834
```

```
  Query Match             1.2%;  Score 51;  DB 8;  Length 43982;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-14;
  Matches   51;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        215 CTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAA 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      34623 CTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAA 34573


RESULT 13
AX347002/c
LOCUS       AX347002               15046 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 2073 from Patent WO0200928.
```

```
ACCESSION   AX347002
VERSION     AX347002.1  GI:18494888
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 2073 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .15046
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             0.6%;  Score 26;  DB 6;  Length 15046;
  Best Local Similarity   100.0%;  Pred. No. 0.35;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3349 AAAAAATATATAATAAAAATAATAAA 3374
              ||||||||||||||||||||||||||
Db       9895 AAAAAATATATAATAAAAATAATAAA 9870


RESULT 14
AC144806
LOCUS       AC144806               124550 bp    DNA     linear   PLN 18-SEP-2003
DEFINITION  Medicago truncatula clone mth2-34h22, complete sequence.
ACCESSION   AC144806
VERSION     AC144806.10  GI:34556237
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 124550)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-34h22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 124550)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 124550)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 124550)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 124550)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Sep 10, 2003 this sequence version replaced gi:34536751.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .124550
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-34h22"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

  Query Match             0.6%;  Score 26;  DB 8;  Length 124550;
  Best Local Similarity   100.0%;  Pred. No. 0.33;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        501 ATATTGTGAAAAAAAAAAAAAAAAGT 526
              ||||||||||||||||||||||||||
Db      69422 ATATTGTGAAAAAAAAAAAAAAAAGT 69447


RESULT 15
AC142222/c
LOCUS       AC142222               125992 bp    DNA     linear   HTG 31-MAY-2003
DEFINITION  Medicago truncatula clone mth2-27d20, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC142222
VERSION     AC142222.7  GI:31249775
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 125992)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-27d20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 125992)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 125992)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 31, 2003 this sequence version replaced gi:30698588.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4334: contig of 4334 bp in length
```

```
            *     4335     4434: gap of unknown length
            *     4435    29695: contig of 25261 bp in length
            *    29696    29795: gap of unknown length
            *    29796    69537: contig of 39742 bp in length
            *    69538    69637: gap of unknown length
            *    69638   125992: contig of 56355 bp in length.
FEATURES             Location/Qualifiers
     source          1. .125992
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-27d20"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

  Query Match              0.6%;  Score 26;  DB 2;  Length 125992;
  Best Local Similarity  100.0%;  Pred. No. 0.33;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        501 ATATTGTGAAAAAAAAAAAAAAAAGT 526
              ||||||||||||||||||||||||||
Db      62571 ATATTGTGAAAAAAAAAAAAAAAAGT 62546


Search completed: July 31, 2004, 15:50:55
Job time : 16512 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 08:20:44 ; Search time 1565 Seconds
(without alignments)
11979.104 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatccacagtgaataaat.....tttcatgggttacgagtcca 4413

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

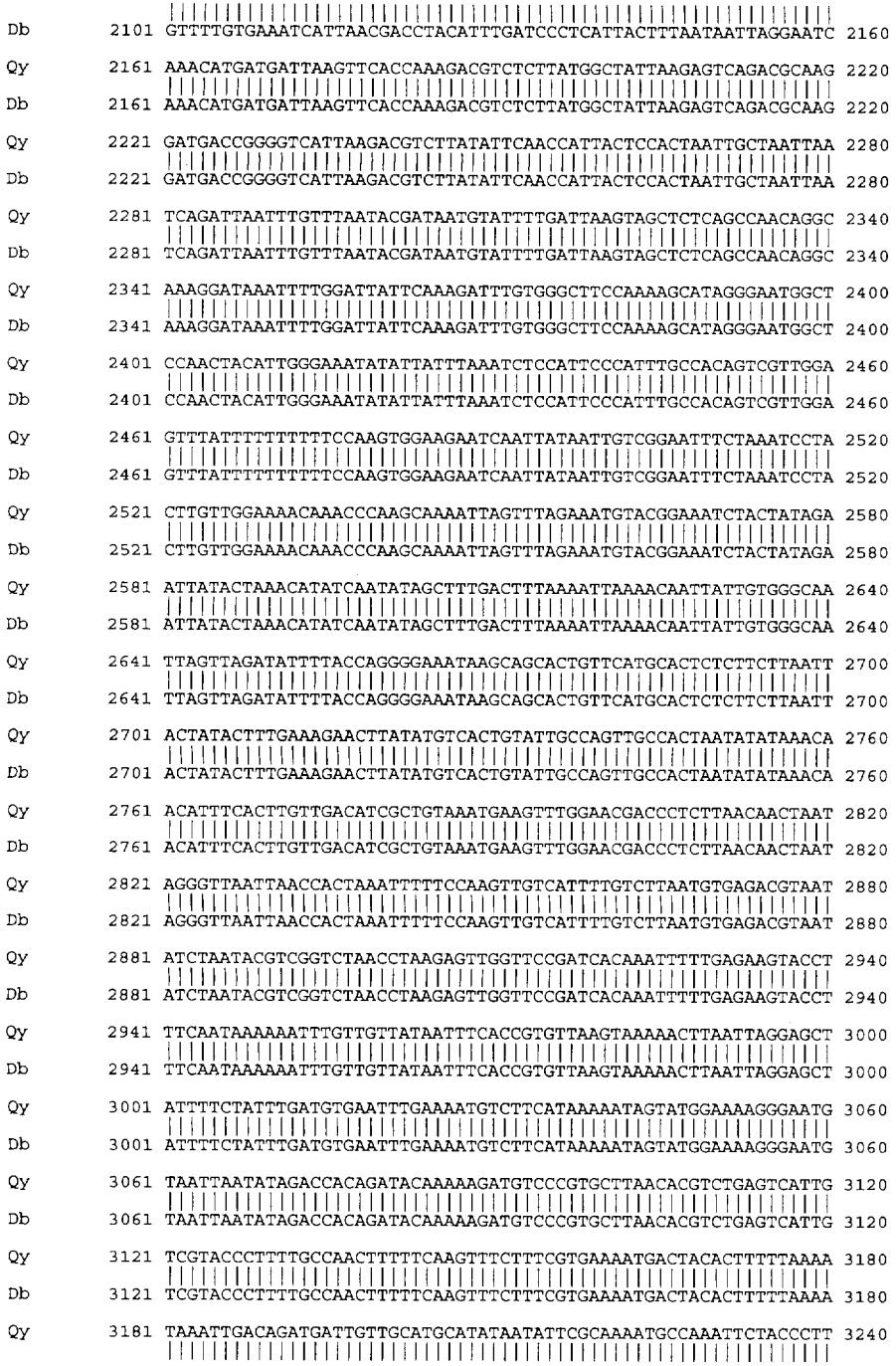
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4413	100.0	4413	4	AAF25318	Aaf25318 Nucleotid
	2	2154	48.8	4309	4	AAF25319	Aaf25319 Nucleotid
	3	2149	48.7	2149	4	AAF25316	Aaf25316 Nucleotid
	4	2149	48.7	4280	4	AAF25317	Aaf25317 Nucleotid
	5	80	1.8	598	3	AAC35408	Aac35408 Arabidops
c	6	26	0.6	15046	6	ABL34100	Abl34100 Human imm
	7	25	0.6	243	5	ABV20030	Abv20030 Human pro
	8	25	0.6	364	5	ABV49789	Abv49789 Human pro
	9	25	0.6	2033	2	AAV40523	Aav40523 Homo sapi
	10	24	0.5	388	8	ACH21196	Ach21196 Human adu
c	11	24	0.5	7752	6	ABL70224	Abl70224 Chemicall
	12	24	0.5	8964	6	ABL33535	Abl33535 Human imm
	13	24	0.5	8964	6	ABK31373	Abk31373 Signal tr
	14	24	0.5	8964	6	ABL70582	Abl70582 Chemicall
	15	24	0.5	8964	6	AAS61277	Aas61277 Human gen
c	16	23	0.5	326	4	AAS38028	Aas38028 Novel hum
	17	23	0.5	367	2	AAX84366	Aax84366 Drosophil
c	18	23	0.5	390	9	ADB55118	Adb55118 Toxicity-
c	19	23	0.5	390	9	ADB49649	Adb49649 Primary r
	20	23	0.5	399	4	AAI88708	Aai88708 Human pol
c	21	23	0.5	458	4	AAL35547	Aal35547 Human mus
c	22	23	0.5	458	7	ABX58535	Abx58535 cDNA enco
	23	23	0.5	976	4	AAL37396	Aal37396 Human mus
	24	23	0.5	976	7	ABX60384	Abx60384 cDNA enco
	25	23	0.5	2292	5	AAS88155	Aas88155 DNA encod
	26	23	0.5	2501	6	ABZ82595	Abz82595 Human sec
	27	23	0.5	2532	9	ADB62172	Adb62172 Human cDN
	28	23	0.5	4636	2	AAZ30710	Aaz30710 Rat neuro
	29	23	0.5	4654	4	AAS46781	Aas46781 Tumour su
	30	23	0.5	4654	6	ABL34223	Abl34223 Human imm
	31	23	0.5	6311	6	ABL33963	Abl33963 Human imm
	32	23	0.5	110000	3	AAF22303_2	Continuation (3 of
	33	23	0.5	203654	7	ABX16034	Abx16034 Human gen
	34	22	0.5	181	4	AAI88838	Aai88838 Human pol
	35	22	0.5	182	6	ABL77052	Abl77052 Human ova
c	36	22	0.5	241	5	ABV07853	Abv07853 Human pro
	37	22	0.5	256	6	ABL80025	Abl80025 Human ova
	38	22	0.5	294	8	ACH16395	Ach16395 Human adu
c	39	22	0.5	373	5	ABV37777	Abv37777 Human pro
	40	22	0.5	375	4	AAL22740	Aal22740 Human bre
	41	22	0.5	391	6	ABV99309	Abv99309 Marine sn
c	42	22	0.5	396	4	AAF94948	Aaf94948 Human ova
c	43	22	0.5	396	6	ABL48898	Abl48898 Ovarian c
c	44	22	0.5	396	6	ABT03215	Abt03215 Human ova
	45	22	0.5	397	4	AAL13874	Aal13874 Human bre

ALIGNMENTS

RESULT 1
AAF25318
ID AAF25318 standard; DNA; 4413 BP.
XX
AC AAF25318;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of an Arabidopsis thaliana genomic fragment.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR001768.
XX
PR 25-JUN-1999; 99FR-00008185.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful for
PT controlling expression of pest or drought resistance genes, and related
PT transgenic plants.
XX
PS Claim 3; Page 71-73; 92pp; French.
XX
CC The present sequence represents a genomic fragment of Arabidopsis
CC thaliana. The sequence contains a plant promoter that directs expression
CC of a selected sequence in root cells at all stages of development of a
CC plant. The plant promoter is used to control expression of genes in a
CC root-specific manner, especially genes that provide resistance to
CC parasites, pests (nematodes or fungi), water and salt stress, or alter
CC sugar content or nitrogen transport. Fragments of the promoter are useful
CC as probes or primers to detect or amplify at least part of the promoter
XX
SQ Sequence 4413 BP; 1427 A; 731 C; 702 G; 1553 T; 0 U; 0 Other;

Query Match 100.0%; Score 4413; DB 4; Length 4413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60

Qy         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120

Qy        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180

Qy        181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240

Qy        241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300

Qy        301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360

Qy        361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420

Qy        421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480

Qy        481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540

Qy        541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600

Qy        601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660

Qy        661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720

Qy        721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780

Qy        781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 840

Qy        841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900

Qy        901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960

Qy        961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020

Qy       1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080

Qy       1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140

Qy       1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200

Qy       1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260

Qy       1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320

Qy       1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380

Qy       1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440

Qy       1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500

Qy       1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560

Qy       1561 TTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTGCAACAACAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620

Qy       1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680

Qy       1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740

Qy       1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800

Qy       1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860

Qy       1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920

Qy       1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980

Qy       1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040

Qy       2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100

Qy       2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160
```


Db 2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160
Qy 2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
Db 2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
Qy 2221 GATGACCGGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280
Db 2221 GATGACCGGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280
Qy 2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340
Db 2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340
Qy 2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400
Db 2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400
Qy 2401 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 2460
Db 2401 CCAACTACATTGGGAAATATATTATTTAAATCTCCATTCCCATTTGCCACAGTCGTTGGA 2460
Qy 2461 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 2520
Db 2461 GTTTATTTTTTTTTTCCAAGTGGAAGAATCAATTATAATTGTCGGAATTTCTAAATCCTA 2520
Qy 2521 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 2580
Db 2521 CTTGTTGGAAAACAAACCCAAGCAAAATTAGTTTAGAAATGTACGGAAATCTACTATAGA 2580
Qy 2581 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 2640
Db 2581 ATTATACTAAACATATCAATATAGCTTTGACTTTAAAATTAAAACAATTATTGTGGGCAA 2640
Qy 2641 TTAGTTAGATATTTTACCAGGGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 2700
Db 2641 TTAGTTAGATATTTTACCAGGGGAAATAAGCAGCACTGTTCATGCACTCTCTTCTTAATT 2700
Qy 2701 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 2760
Db 2701 ACTATACTTTGAAAGAACTTATATGTCACTGTATTGCCAGTTGCCACTAATATATAAACA 2760
Qy 2761 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 2820
Db 2761 ACATTTCACTTGTTGACATCGCTGTAAATGAAGTTTGGAACGACCCTCTTAACAACTAAT 2820
Qy 2821 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
Db 2821 AGGGTTAATTAACCACTAAATTTTTCCAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
Qy 2881 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 2940
Db 2881 ATCTAATACGTCGGTCTAACCTAAGAGTTGGTTCCGATCACAAATTTTTGAGAAGTACCT 2940
Qy 2941 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 3000
Db 2941 TTCAATAAAAAATTTGTTGTTATAATTTCACCGTGTTAAGTAAAAACTTAATTAGGAGCT 3000
Qy 3001 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 3060
Db 3001 ATTTTCTATTTGATGTGAATTTGAAAATGTCTTCATAAAAATAGTATGGAAAAGGGAATG 3060
Qy 3061 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 3120
Db 3061 TAATTAATATAGACCACAGATACAAAAAGATGTCCCGTGCTTAACACGTCTGAGTCATTG 3120
Qy 3121 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 3180
Db 3121 TCGTACCCTTTTGCCAACTTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTTAAAA 3180
Qy 3181 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 3240
Db 3181 TAAATTGACAGATGATTGTTGCATGCATATAATATTCGCAAAATGCCAAATTCTACCCTT 3240
Qy 3241 AACCAAATAATGGGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTTG 3300
Db 3241 AACCAAATAATGGGTAATGGATATAAAATAGTTAACATAAACAAAAATTTGGAAATTTTG 3300
Qy 3301 AAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATATA 3360
Db 3301 AAAAATAGAGAGTAAATTGATTTTTTTTAAAAAGTTTGAATTGAAGTGAAAAAATATATA 3360
Qy 3361 ATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAACT 3420
Db 3361 ATAAAAATAATAAACCTGTAGTTTGATATATATAGTTAGATAGCTCAAGTTTGAGTAACT 3420
Qy 3421 TGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATCCC 3480
Db 3421 TGAAGTCTTGAATTACTTTATATGTTTTTCTCACAGATTATATTATTTTCGTTCTATCCC 3480
Qy 3481 AAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTTAAGAAGTCGGTGTATAAT 3540
Db 3481 AAGAATTGGGATAATATTCTCTATATTCGAGGCCTCTCTCTTTAAGAAGTCGGTGTATAAT 3540
Qy 3541 CTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTTTC 3600
Db 3541 CTTAAGCCCTTACTTGACACAAGGCCTCTATTAAAAAGCCCAATAATAATTTTTCTTTTC 3600
Qy 3601 AAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGGCG 3660
Db 3601 AAAGCCCAACACGTCAAGAGGAGAAAGAAGTGCGTTTGCGTTTGGATTGAAAACGTGGCG 3660
Qy 3661 GGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCTCT 3720
Db 3661 GGCGGTTGGAAACTGAATCTTAAACCCTTCACTCACATTCATCTTCACCGTCAAATCTCT 3720
Qy 3721 GAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
Db 3721 GAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
Qy 3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACATTT 3840
Db 3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAACACTTCAGCGCCACATTT 3840
Qy 3841 GGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGATTT 3900
Db 3841 GGGTTTTACGATTAGGGTTTCAATAGATCGAATCGATTCATCTCTCTTGGTATTAGATTT 3900
Qy 3901 CATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCGGG 3960
Db 3901 CATCAAATTCAAATCTATCGTCACCTAAAAATCTCTCTTTCGGTATGTAGTATTTCCGGG 3960
Qy 3961 GTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
Db 3961 GTTGTTTCAAAGTCGTCTTATTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
Qy 4021 CTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGTCT 4080
Db 4021 CTATGTAGAAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATAGTTCAATGTCT 4080
Qy 4081 ACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTTGT 4140
Db 4081 ACATTATGAGATTGCCATGATATTATAGTCAAACGATTCTCCTAAAGCGTTTACTTTTGT 4140
Qy 4141 GGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGAAG 4200
Db 4141 GGAGCATTCTTTGTTAGTTCTATGCAAATAAAGTTCTAGGAATGATAATTCTTAAGGAAG 4200
Qy 4201 CATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTTTT 4260
Db 4201 CATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTTGCAATTTGCTTTT 4260
Qy 4261 AGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGTTT 4320
Db 4261 AGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTGTGAAGCAAAGTTT 4320

```
Qy       4321 GATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTATTT 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATCAATTTATGTATTT 4380

Qy       4381 TCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 4413
             |||||||||||||||||||||||||||||||||
Db       4381 TCTTTGTCTACAGTTTCATGGGTTACGAGTCCA 4413


RESULT 2
AAF25319
ID   AAF25319 standard; DNA; 4309 BP.
XX
AC   AAF25319;
XX
DT   30-APR-2001  (first entry)
XX
DE   Nucleotide sequence of a plant promoter insert in plasmid pBin19.
XX
KW   Plant promoter; root cell; root-specific expression; parasite resistance;
KW   nematode resistance; fungal resistance; water stress; salt stress;
KW   sugar content; nitrogen transport; ss.
XX
OS   Synthetic.
OS   Arabidopsis thaliana.
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   promoter        30. .2178
FT                   /*tag= a
FT   misc_feature    2179. .4309
FT                   /*tag= b
FT                   /note= "GUS coding region"
XX
PN   WO200100833-A1.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-FR001768.
XX
PR   25-JUN-1999;   99FR-00008185.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Hoffmann B,  Mollier P,  Pelletier G;
XX
DR   WPI; 2001-102893/11.
XX
PT   New constitutive plant promoter active specifically in roots, useful for
PT   controlling expression of pest or drought resistance genes, and related
PT   transgenic plants.
XX
PS   Example 3; Page 73-75; 92pp; French.
XX
CC   The present sequence represents a plant promoter insert in plasmid
CC   pBin19. The plant promoter directs expression of a selected sequence in
CC   root cells at all stages of development of a plant. The plant promoter is
CC   used to control expression of genes in a root-specific manner, especially
CC   genes that provide resistance to parasites, pests (nematodes or fungi),
CC   water and salt stress, or alter sugar content or nitrogen transport.
CC   Fragments of the promoter are useful as probes or primers to detect or
CC   amplify at least part of the promoter
XX
SQ   Sequence 4309 BP; 1262 A; 868 C; 915 G; 1264 T; 0 U; 0 Other;

  Query Match             48.8%;  Score 2154;  DB 4;  Length 4309;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2154;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        131 CTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAAC 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 CTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAAC 84

Qy        191 AACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTAC 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 AACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTAC 144

Qy        251 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 204

Qy        311 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 264

Qy        371 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 324

Qy        431 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 384

Qy        491 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 444

Qy        551 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 504

Qy        611 TTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 TTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 564

Qy        671 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 624

Qy        731 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 684

Qy        791 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACG 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACG 744

Qy        851 TATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 TATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 804

Qy        911 TGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCA 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCA 864

Qy        971 TATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTA 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 TATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTA 924

Qy       1031 TTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTC 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 TTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTC 984

Qy       1091 TACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTG 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 TACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTG 1044

Qy       1151 AAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGA 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 AAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGA 1104

Qy       1211 ATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACA 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 ATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACA 1164
```

```
Qy       1271 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1330
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1165 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1224

Qy       1331 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1225 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1284

Qy       1391 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTT 1450
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1285 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTT 1344

Qy       1451 CTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCAT 1510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1345 CTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCAT 1404

Qy       1511 AGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAA 1570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1405 AGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAA 1464

Qy       1571 CAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGT 1630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1465 CAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGT 1524

Qy       1631 CGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATA 1690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1525 CGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATA 1584

Qy       1691 ATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 ATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1644

Qy       1751 TCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTA 1810
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1645 TCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTA 1704

Qy       1811 ATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAA 1870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1705 ATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAA 1764

Qy       1871 TGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAA 1930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1765 TGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAA 1824

Qy       1931 CATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTG 1990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1825 CATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTG 1884

Qy       1991 TTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACA 2050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1885 TTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACA 1944

Qy       2051 TTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAA 2110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1945 TTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAA 2004

Qy       2111 ATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATG 2170
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2005 ATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATG 2064

Qy       2171 ATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2230
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2065 ATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2124

Qy       2231 GTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2125 GTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2178
```

RESULT 3
AAF25316
ID AAF25316 standard; DNA; 2149 BP.
XX
AC AAF25316;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plant promoter for expression in roots.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR001768.
XX
PR 25-JUN-1999; 99FR-00008185.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful for
PT controlling expression of pest or drought resistance genes, and related
PT transgenic plants.
XX
PS Claim 1; Fig 1; 92pp; French.
XX
CC The present sequence represents a plant promoter that directs expression
CC of a selected sequence in root cells at all stages of development of a
CC plant. The plant promoter is used to control expression of genes in a
CC root-specific manner, especially genes that provide resistance to
CC parasites, pests (nematodes or fungi), water and salt stress, or alter
CC sugar content or nitrogen transport. Fragments of the promoter are useful
CC as probes or primers to detect or amplify at least part of the promoter
XX
SQ Sequence 2149 BP; 693 A; 365 C; 335 G; 756 T; 0 U; 0 Other;

Query Match 48.7%; Score 2149; DB 4; Length 2149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy        196 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420
```

```
Qy       556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615
Db       421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy       616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675
Db       481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy       676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735
Db       541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy       736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795
Db       601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy       796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855
Db       661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy       856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915
Db       721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy       916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975
Db       781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy       976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035
Db       841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy      1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095
Db       901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy      1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155
Db       961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy      1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215
Db      1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy      1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275
Db      1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy      1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335
Db      1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy      1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395
Db      1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy      1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455
Db      1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy      1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515
Db      1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy      1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575
Db      1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy      1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635
Db      1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy      1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695
Db      1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy      1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
Db      1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy      1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815
Db      1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy      1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875
Db      1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy      1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935
Db      1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy      1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995
Db      1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy      1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
Db      1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy      2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115
Db      1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy      2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175
Db      1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy      2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235
Db      2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy      2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
Db      2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

RESULT 4
AAF25317
ID AAF25317 standard; DNA; 4280 BP.
XX
AC AAF25317;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a root promoter and the GSU gene coding region.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; GUS gene; ss.
XX
OS Synthetic.
OS Arabidopsis thaliana.
OS Unidentified.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR001768.
XX
PR 25-JUN-1999; 99FR-00008185.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX

PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful for
PT controlling expression of pest or drought resistance genes, and related
PT transgenic plants.
XX
PS Claim 5; Page 69-71; 92pp; French.
XX
CC The present sequence represents a plant promoter linked to a GUS coding
CC region. The plant promoter directs expression of a selected sequence in
CC root cells at all stages of development of a plant. The plant promoter is
CC used to control expression of genes in a root-specific manner, especially
CC genes that provide resistance to parasites, pests (nematodes or fungi),
CC water and salt stress, or alter sugar content or nitrogen transport.
CC Fragments of the promoter are useful as probes or primers to detect or
CC amplify at least part of the promoter
XX
SQ Sequence 4280 BP; 1256 A; 859 C; 907 G; 1258 T; 0 U; 0 Other;

Query Match 48.7%; Score 2149; DB 4; Length 4280;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy      196 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy      256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy      316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy      376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy      436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy      496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy      556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy      616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy      676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy      736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy      796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy      856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy      916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy      976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy     1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy     1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy     1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy     1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy     1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy     1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy     1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy     1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy     1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy     1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy     1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy     1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy     1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy     1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy     1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy     1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

RESULT 5
AAC35408
ID AAC35408 standard; DNA; 598 BP.
XX
AC AAC35408;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10084.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.8%; Score 80; DB 3; Length 598;
Best Local Similarity 100.0%; Pred. No. 5.2e-20;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3727 GAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCAGCCAGT 3786
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCAGCCAGT 60

Qy 3787 CTCTGCTTCGTAATGTTACG 3806
||||||||||||||||||||
Db 61 CTCTGCTTCGTAATGTTACG 80

RESULT 6
ABL34100/c
ID ABL34100 standard; DNA; 15046 BP.
XX
AC ABL34100;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2073.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 2073; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 15046 BP; 4162 A; 255 C; 3463 G; 7166 T; 0 U; 0 Other;

```
  Query Match             0.6%;  Score 26;  DB 6;  Length 15046;
  Best Local Similarity   100.0%;  Pred. No. 5.2;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3349 AAAAAATATATAATAAAAATAATAAA 3374
              ||||||||||||||||||||||||||
Db       9895 AAAAAATATATAATAAAAATAATAAA 9870
```

RESULT 7
ABV20030
ID ABV20030 standard; cDNA; 243 BP.
XX
AC ABV20030;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 20021.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 3266; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 243 BP; 126 A; 20 C; 39 G; 58 T; 0 U; 0 Other;

```
  Query Match             0.6%;  Score 25;  DB 5;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        500 TATATTGTGAAAAAAAAAAAAAAAA 524
              |||||||||||||||||||||||||
Db        169 TATATTGTGAAAAAAAAAAAAAAAA 193
```

RESULT 8
ABV49789
ID ABV49789 standard; cDNA; 364 BP.
XX
AC ABV49789;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 49780.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 9710; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 364 BP; 146 A; 72 C; 69 G; 76 T; 0 U; 1 Other;

```
  Query Match             0.6%;  Score 25;  DB 5;  Length 364;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        500 TATATTGTGAAAAAAAAAAAAAAAA 524
              |||||||||||||||||||||||||
Db        207 TATATTGTGAAAAAAAAAAAAAAAA 231
```

RESULT 9
AAV40523
ID AAV40523 standard; cDNA; 2033 BP.
XX
AC AAV40523;
XX
DT 27-OCT-1998 (first entry)
XX

```
DE   Homo sapiens CN483_2 clone secreted protein coding region.
XX
KW   secreted protein; CN483_2; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             112. .1215
FT                   /*tag= a
FT                   /note= "secreted protein"
XX
PN   WO9830695-A2.
XX
PD   16-JUL-1998.
XX
PF   09-JAN-1998;   98WO-US000543.
XX
PR   09-JAN-1997;   97US-00780814.
PR   08-JAN-1998;   98US-00004684.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Racie LA,  Merberg D,  Treacy M;
PI   Spaulding V,  Agostino MJ;
XX
DR   WPI; 1998-413686/35.
DR   P-PSDB; AAW29662.
XX
PT   New isolated nucleic acids and secreted proteins - obtained from human
PT   adult ovary, human foetal kidney, human foetal brain and human adult
PT   brain cDNA libraries.
XX
PS   Claim 22; Page 73-74; 113pp; English.
XX
CC   The sequence is that encoding a novel, isolated secreted protein
XX
SQ   Sequence 2033 BP; 611 A; 386 C; 517 G; 513 T; 0 U; 6 Other;

  Query Match             0.6%;  Score 25;  DB 2;  Length 2033;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        500 TATATTGTGAAAAAAAAAAAAAAAA 524
              ||||||||||||||||||||||||||
Db       2009 TATATTGTGAAAAAAAAAAAAAAAA 2033


RESULT 10
ACH21196
ID   ACH21196 standard; cDNA; 388 BP.
XX
AC   ACH21196;
XX
DT   13-OCT-2003  (first entry)
XX
DE   Human adult liver cDNA #808.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT,  Labat I,  Stache-Crain B,  Dickson MC,  Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 8408; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 388 BP; 115 A; 89 C; 95 G; 89 T; 0 U; 0 Other;

  Query Match             0.5%;  Score 24;  DB 8;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 TATTGTGAAAAAAAAAAAAAAAAG 525
              ||||||||||||||||||||||||
Db        260 TATTGTGAAAAAAAAAAAAAAAAG 283


RESULT 11
ABL70224/c
ID   ABL70224 standard; DNA; 7752 BP.
XX
AC   ABL70224;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated cell signalling DNA sequence complementary to#57.
XX
KW   Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW   tumour; cytostatic; ds.
XX
OS   Unidentified.
XX
PN   WO200202807-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007471.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154758/20.
XX
```

PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX
PS Claim 1; SEQ ID NO 114; 24pp + Sequence Listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 7752 BP; 1847 A; 113 C; 1696 G; 4096 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 7752;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3349 AAAAAATATATAATAAAAATAATA 3372
||||||||||||||||||||||||
Db 4466 AAAAAATATATAATAAAAATAATA 4443

RESULT 12
ABL33535
ID ABL33535 standard; DNA; 8964 BP.
XX
AC ABL33535;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1508.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1508; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3350 AAAAATATATAATAAAAATAATAA 3373
||||||||||||||||||||||||
Db 7648 AAAAATATATAATAAAAATAATAA 7671

RESULT 13
ABK31373
ID ABK31373 standard; DNA; 8964 BP.
XX
AC ABK31373;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified complementary DNA #108.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 216; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are

CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3350 AAAAATATATAATAAAAATAATAA 3373
||||||||||||||||||||||||
Db 7648 AAAAATATATAATAAAAATAATAA 7671

RESULT 14
ABL70582
ID ABL70582 standard; DNA; 8964 BP.
XX
AC ABL70582;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence complementary to#236.
XX
KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007471.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX
PS Claim 1; SEQ ID NO 472; 24pp + Sequence Listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3350 AAAAATATATAATAAAAATAATAA 3373
||||||||||||||||||||||||
Db 7648 AAAAATATATAATAAAAATAATAA 7671

RESULT 15
AAS61277
ID AAS61277 standard; DNA; 8964 BP.
XX
AC AAS61277;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #232.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003968.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.
XX
PS Disclosure; SEQ ID NO 238; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3350 AAAAATATATAATAAAAATAATAA 3373
              ||||||||||||||||||||||||
Db       7648 AAAAATATATAATAAAAATAATAA 7671
```

Search completed: July 31, 2004, 11:15:39
Job time : 1574 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 10:32:54 ; Search time 285 Seconds
(without alignments)
8592.976 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatccacagtgaataaat.....tttcatgggttacgagtcca 4413

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23	0.5	367	4	US-09-219-983A-13	Sequence 13, Appl
c	2	22	0.5	396	4	US-09-640-173-139	Sequence 139, App
c	3	22	0.5	396	4	US-09-713-550-139	Sequence 139, App
	4	22	0.5	449	4	US-09-621-976-18967	Sequence 18967, A
	5	22	0.5	1790	4	US-09-866-028-31	Sequence 31, Appl
	6	22	0.5	1866	4	US-09-224-048A-1	Sequence 1, Appli
c	7	22	0.5	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	8	21	0.5	35	1	US-08-126-594-20	Sequence 20, Appl
c	9	21	0.5	35	1	US-08-465-811A-20	Sequence 20, Appl
c	10	21	0.5	35	2	US-08-619-542B-20	Sequence 20, Appl
	11	21	0.5	326	4	US-09-629-645A-17	Sequence 17, Appl
	12	21	0.5	454	4	US-09-621-976-9423	Sequence 9423, Ap
	13	21	0.5	780	4	US-09-328-475C-145	Sequence 145, App
	14	21	0.5	1198	3	US-08-849-751-3	Sequence 3, Appli
	15	21	0.5	1198	4	US-09-478-816-3	Sequence 3, Appli
	16	21	0.5	1267	1	US-08-140-797-1	Sequence 1, Appli
	17	21	0.5	1267	1	US-08-486-670A-1	Sequence 1, Appli
	18	21	0.5	1292	4	US-09-620-312D-868	Sequence 868, App
c	19	21	0.5	1415	1	US-08-236-918A-7	Sequence 7, Appli
c	20	21	0.5	1415	4	US-09-150-864A-7	Sequence 7, Appli
	21	21	0.5	1586	4	US-09-178-093B-3	Sequence 3, Appli
c	22	21	0.5	1601	4	US-09-220-132-182	Sequence 182, App
	23	21	0.5	1688	2	US-08-439-814-2	Sequence 2, Appli
	24	21	0.5	1716	2	US-08-954-333-9	Sequence 9, Appli
	25	21	0.5	1779	3	US-09-323-427-1	Sequence 1, Appli
c	26	21	0.5	1779	3	US-09-323-427-2	Sequence 2, Appli
	27	21	0.5	1779	4	US-09-812-642-1	Sequence 1, Appli
c	28	21	0.5	1779	4	US-09-812-642-2	Sequence 2, Appli
	29	21	0.5	1786	1	US-07-920-430-19	Sequence 19, Appl
	30	21	0.5	1786	1	US-08-066-299-9	Sequence 9, Appli
	31	21	0.5	1786	1	US-08-265-047-1	Sequence 1, Appli
	32	21	0.5	1786	1	US-08-251-464-19	Sequence 19, Appl
	33	21	0.5	1786	4	US-08-926-522-20	Sequence 20, Appl
	34	21	0.5	1786	5	PCT-US92-01364-19	Sequence 19, Appl
	35	21	0.5	2050	4	US-09-688-188B-28	Sequence 28, Appl
	36	21	0.5	2050	4	US-09-291-417D-28	Sequence 28, Appl
	37	21	0.5	2090	2	US-08-439-814-1	Sequence 1, Appli
c	38	21	0.5	2274	4	US-09-220-132-188	Sequence 188, App
	39	21	0.5	2475	4	US-09-220-132-179	Sequence 179, App
	40	21	0.5	2638	4	US-09-228-986-8	Sequence 8, Appli
	41	21	0.5	2720	2	US-08-007-107-3	Sequence 3, Appli
	42	21	0.5	2738	4	US-09-554-726A-9	Sequence 9, Appli
	43	21	0.5	3282	4	US-09-489-847-52	Sequence 52, Appl
	44	21	0.5	3347	4	US-09-702-705-318	Sequence 318, App
	45	21	0.5	3347	4	US-09-736-457-318	Sequence 318, App

ALIGNMENTS

RESULT 1
US-09-219-983A-13
; Sequence 13, Application US/09219983A
; Patent No. 6380159
; GENERAL INFORMATION:
; APPLICANT: Wolfner, Mariana
; APPLICANT: Lung, Oliver
; APPLICANT: Tram, Khanh-Uyen
; TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
; TITLE OF INVENTION: MELANOGASTER
; FILE REFERENCE: 19603/1791
; CURRENT APPLICATION NUMBER: US/09/219,983A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/071,315
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 367
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-219-983A-13

Query Match 0.5%; Score 23; DB 4; Length 367;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 502 TATTGTGAAAAAAAAAAAAAAAA 524
||||||||||||||||||||||
Db 336 TATTGTGAAAAAAAAAAAAAAAA 358

RESULT 2
US-09-640-173-139/c
; Sequence 139, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 139
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(396)
;   OTHER INFORMATION: n = A,T,C or G
US-09-640-173-139

  Query Match             0.5%;  Score 22;  DB 4;  Length 396;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 TTGTGAAAAAAAAAAAAAAAAG 525
              ||||||||||||||||||||||
Db         27 TTGTGAAAAAAAAAAAAAAAAG 6


RESULT 3
US-09-713-550-139/c
; Sequence 139, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Stolk, John A.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE:  210121.484C4
;  CURRENT APPLICATION NUMBER: US/09/713,550
;  CURRENT FILING DATE:  2000-11-14
;  NUMBER OF SEQ ID NOS:  205
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 139
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(396)
;   OTHER INFORMATION: n = A,T,C or G
US-09-713-550-139

  Query Match             0.5%;  Score 22;  DB 4;  Length 396;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 TTGTGAAAAAAAAAAAAAAAAG 525
              ||||||||||||||||||||||
Db         27 TTGTGAAAAAAAAAAAAAAAAG 6


RESULT 4
US-09-621-976-18967
; Sequence 18967, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 18967
;   LENGTH: 449
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-18967

  Query Match             0.5%;  Score 22;  DB 4;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 TTGTGAAAAAAAAAAAAAAAAG 525
              ||||||||||||||||||||||
Db        428 TTGTGAAAAAAAAAAAAAAAAG 449


RESULT 5
US-09-866-028-31
; Sequence 31, Application US/09866028
; Patent No. 6642360
; GENERAL INFORMATION:
;  APPLICANT: Baker, Kevin
;  APPLICANT:  Botstein,David
;  APPLICANT:  Eaton,Dan
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul
;  APPLICANT:  Grimaldi,Christopher
;  APPLICANT:  Gurney,Austin
;  APPLICANT:  Hillan,Kenneth
;  APPLICANT:  Kljavin,Ivar
;  APPLICANT:  Napier,Mary
;  APPLICANT:  Roy,Margaret
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Wood,William
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P2548P1C1
;  CURRENT APPLICATION NUMBER: US/09/866,028
;  CURRENT FILING DATE:  2001-05-25
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 31
;   LENGTH: 1790
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-866-028-31

  Query Match             0.5%;  Score 22;  DB 4;  Length 1790;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        503 ATTGTGAAAAAAAAAAAAAAAA 524
              ||||||||||||||||||||||
Db       1721 ATTGTGAAAAAAAAAAAAAAAA 1742


RESULT 6
US-09-224-048A-1
; Sequence 1, Application US/09224048A
; Patent No. 6387366
; GENERAL INFORMATION:
;  APPLICANT: Hurwitz, David R.
;  APPLICANT:  Cherington, Van
;  APPLICANT:  Galanopoulos, Theofanis
;  APPLICANT:  Levine, Peter H.
;  APPLICANT:  Greenberger, Joel S.
;  TITLE OF INVENTION: METHOD FOR REDUCING ADVERSE SIDE EFFECTS ASSOCIATED
;  TITLE OF INVENTION:  WITH BONE MARROW CELL TRANSPLANTATION
;  FILE REFERENCE: 07787/007001
;  CURRENT APPLICATION NUMBER: US/09/224,048A
;  CURRENT FILING DATE:  1998-12-31
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1866
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (111)...(986)
US-09-224-048A-1
```

```
  Query Match             0.5%;  Score 22;  DB 4;  Length 1866;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        503 ATTGTGAAAAAAAAAAAAAAAA 524
              ||||||||||||||||||||||
Db       1834 ATTGTGAAAAAAAAAAAAAAAA 1855


RESULT 7
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             0.5%;  Score 22;  DB 4;  Length 1664976;
  Best Local Similarity   100.0%;  Pred. No. 6.5;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1353 TATTGAAGATGTTTATATATAT 1374
              ||||||||||||||||||||||
Db    1283595 TATTGAAGATGTTTATATATAT 1283574


RESULT 8
US-08-126-594-20/c
; Sequence 20, Application US/08126594
; Patent No. 5482845
;  GENERAL INFORMATION:
;    APPLICANT:  Soares, M. Bento
;    APPLICANT:  Efstratiadis, Algiris
;    TITLE OF INVENTION:  METHOD FOR CONSTRUCTION OF NORMALIZED
;    TITLE OF INVENTION:  cDNA LIBRARIES
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White, c/o Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/126,594
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  42840/JPW/AKC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-126-594-20

  Query Match             0.5%;  Score 21;  DB 1;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        509 AAAAAAAAAAAAAAAAAGTGGT 529
              |||||||||||||||||||||
Db         34 AAAAAAAAAAAAAAAAAGTGGT 14


RESULT 9
US-08-465-811A-20/c
; Sequence 20, Application US/08465811A
; Patent No. 5637685
;  GENERAL INFORMATION:
;    APPLICANT:  Soares, M. Bento
;    APPLICANT:  Efstratiadis, Argiris
;    TITLE OF INVENTION:  METHOD FOR CONSTRUCTION OF
;    TITLE OF INVENTION:  NORMALIZED cDNA LIBRARIES
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White, c/o Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,811A
;      FILING DATE:  June 6, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  42840/JPW/AKC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-465-811A-20

  Query Match             0.5%;  Score 21;  DB 1;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        509 AAAAAAAAAAAAAAAAAGTGGT 529
              |||||||||||||||||||||
Db         34 AAAAAAAAAAAAAAAAAGTGGT 14


RESULT 10
US-08-619-542B-20/c
; Sequence 20, Application US/08619542B
; Patent No. 5830662
;  GENERAL INFORMATION:
;    APPLICANT:  The Trustees of Columbia University in the City
;    APPLICANT:  of New York
;    TITLE OF INVENTION:  METHOD FOR CONSTRUCTION OF NORMALIZED
;    TITLE OF INVENTION:  cDNA LIBRARIES
;    NUMBER OF SEQUENCES:  78
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,542B
;      FILING DATE:  June 21, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  42840-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-619-542B-20

  Query Match             0.5%;  Score 21;  DB 2;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        509 AAAAAAAAAAAAAAAAAGTGGT 529
              |||||||||||||||||||||
Db         34 AAAAAAAAAAAAAAAAAGTGGT 14


RESULT 11
US-09-629-645A-17
; Sequence 17, Application US/09629645A
; Patent No. 6365354
; GENERAL INFORMATION:
;  APPLICANT: C. Frank Bennett
;  APPLICANT:  Jacqueline Wyatt
;  TITLE OF INVENTION: ANTISENSE MODULATION OF LYSOPHOSPHOLIPASE I EXPRESSION
;  FILE REFERENCE: RTS-0137
;  CURRENT APPLICATION NUMBER: US/09/629,645A
;  CURRENT FILING DATE:  2000-07-31
;  NUMBER OF SEQ ID NOS: 164
; SEQ ID NO 17
;   LENGTH: 326
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 99
;   OTHER INFORMATION: unknown
US-09-629-645A-17

  Query Match             0.5%;  Score 21;  DB 4;  Length 326;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 TTGTGAAAAAAAAAAAAAAAA 524
              |||||||||||||||||||||
Db        275 TTGTGAAAAAAAAAAAAAAAA 295


RESULT 12
US-09-621-976-9423
; Sequence 9423, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 9423
;   LENGTH: 454
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 20,386
;   OTHER INFORMATION: n=a, g, c or t
US-09-621-976-9423

  Query Match             0.5%;  Score 21;  DB 4;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 TTGTGAAAAAAAAAAAAAAAA 524
              |||||||||||||||||||||
Db        434 TTGTGAAAAAAAAAAAAAAAA 454


RESULT 13
US-09-328-475C-145
; Sequence 145, Application US/09328475C
; Patent No. 6476207
; GENERAL INFORMATION:
;  APPLICANT: Zhang, Jimmy
;  APPLICANT:  Astel, Jon H.
;  APPLICANT:  Carroll III, Eddie
;  APPLICANT:  Endege, Wilson O.
;  APPLICANT:  Ford, Donna M.
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Steinmann, Kathleen E.
;  TITLE OF INVENTION: GENES AND GENE EXPRESSION PRODUCTS THAT
;  TITLE OF INVENTION:  ARE DIFFERENTIALLY REGULATED IN PROSTATE CANCER
```

```
;  FILE REFERENCE: 1532.002/200130.463
;  CURRENT APPLICATION NUMBER: US/09/328,475C
;  CURRENT FILING DATE:  1999-06-09
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 145
;   LENGTH: 780
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(780)
;   OTHER INFORMATION: n = A,T,C or G
US-09-328-475C-145

  Query Match             0.5%;  Score 21;  DB 4;  Length 780;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 TATTGTGAAAAAAAAAAAAA 522
              |||||||||||||||||||||
Db        314 TATTGTGAAAAAAAAAAAAA 334


RESULT 14
US-08-849-751-3
; Sequence 3, Application US/08849751
; Patent No. 6190890
;  GENERAL INFORMATION:
;    APPLICANT:  VAN DEN BROECK, HENRIETTE C.
;    APPLICANT:  DE GRAAFF, LEENDERT H.
;    APPLICANT:  VISSER, JACOB
;    APPLICANT:  VAN OOYEN, ALBERT J.J.
;    TITLE OF INVENTION:  FUNGAL CELLULASES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW, suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/849,751
;      FILING DATE:  12-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP96/04511
;      FILING DATE:  14-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murashige, Kate H
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  24615-20094.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-887-1500
;      TELEFAX:  202-822-0168
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1198 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;      STRAIN:  N400
;      INDIVIDUAL ISOLATE:  CBS120.49
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  32...1024
;      OTHER INFORMATION:  product="Cellulase"
US-08-849-751-3

  Query Match             0.5%;  Score 21;  DB 3;  Length 1198;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 TTGTGAAAAAAAAAAAAAAAA 524
              ||||||||||||||||||||
Db       1168 TTGTGAAAAAAAAAAAAAAAA 1188


RESULT 15
US-09-478-816-3
; Sequence 3, Application US/09478816
; Patent No. 6306635
;  GENERAL INFORMATION:
;    APPLICANT:  VAN DEN BROECK, HENRIETTE C.
;    APPLICANT:  DE GRAAFF, LEENDERT H.
;    APPLICANT:  VISSER, JACOB
;    APPLICANT:  VAN OOYEN, ALBERT J.J.
;    TITLE OF INVENTION:  FUNGAL CELLULASES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW, suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/478,816
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/849,751
;      FILING DATE:  12-JUN-1997
;      APPLICATION NUMBER:  PCT/EP96/04511
;      FILING DATE:  14-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murashige, Kate H
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  24615-20094.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-887-1500
;      TELEFAX:  202-822-0168
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1198 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;      STRAIN:  N400
;      INDIVIDUAL ISOLATE:  CBS120.49
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  32...1024
```

; OTHER INFORMATION: product="Cellulase"
US-09-478-816-3

Query Match 0.5%; Score 21; DB 4; Length 1198;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 TTGTGAAAAAAAAAAAAAAAA 524
||||||||||||||||||||
Db 1168 TTGTGAAAAAAAAAAAAAAAA 1188

Search completed: July 31, 2004, 18:48:11
Job time : 291 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 15:50:59 ; Search time 1818 Seconds
(without alignments)
11901.847 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatccacagtgaataaaat.....tttcatgggttacgagtcca 4413

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3222919 seqs, 2451570024 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

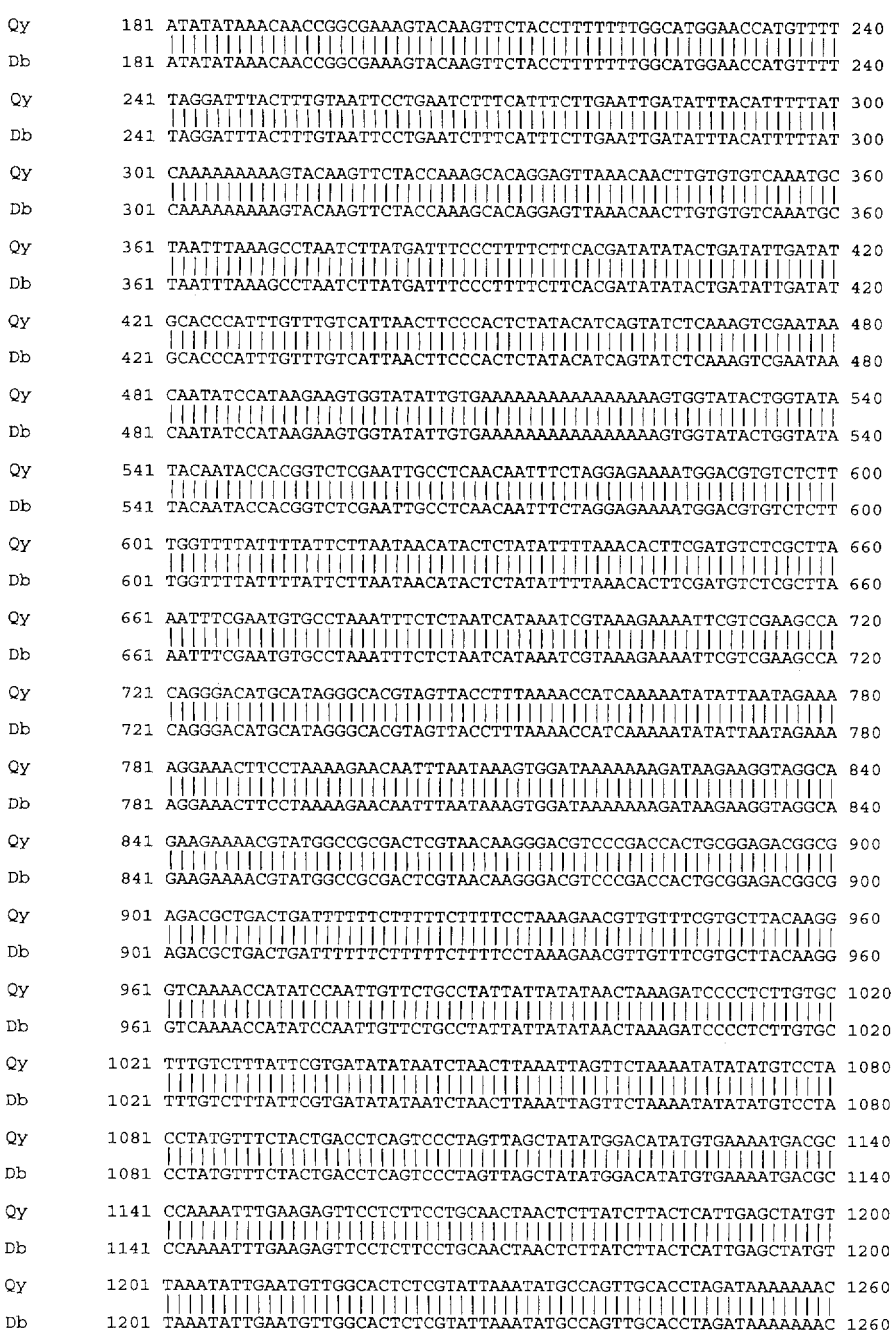
Result No.	Score	% Query Match	Length	DB	ID	Description
1	4413	100.0	4413	15	US-10-027-880-3	Sequence 3, Appli
2	2154	48.8	4309	15	US-10-027-880-4	Sequence 4, Appli
3	2149	48.7	2149	15	US-10-027-880-1	Sequence 1, Appli
4	2149	48.7	4280	15	US-10-027-880-2	Sequence 2, Appli
c 5	26	0.6	15046	15	US-10-311-455-2073	Sequence 2073, Ap
6	25	0.6	68140	13	US-10-087-192-1681	Sequence 1681, Ap
7	24	0.5	388	10	US-09-918-995-8408	Sequence 8408, Ap
c 8	24	0.5	440	13	US-10-424-599-58606	Sequence 58606, A
9	24	0.5	2211	16	US-10-310-154-361	Sequence 361, App
10	24	0.5	8964	13	US-10-221-613-238	Sequence 238, App
11	24	0.5	8964	15	US-10-311-455-1508	Sequence 1508, Ap
c 12	24	0.5	126266	13	US-10-087-192-1132	Sequence 1132, Ap
c 13	24	0.5	132762	10	US-09-954-556-17	Sequence 17, Appl
c 14	23	0.5	326	10	US-09-803-719-1086	Sequence 1086, Ap
15	23	0.5	367	15	US-10-114-774-13	Sequence 13, Appl
c 16	23	0.5	390	16	US-10-388-934-308	Sequence 308, App
c 17	23	0.5	458	9	US-09-764-877-889	Sequence 889, App
c 18	23	0.5	458	16	US-10-242-515-889	Sequence 889, App
c 19	23	0.5	613	16	US-10-388-934-310	Sequence 310, App
c 20	23	0.5	668	13	US-10-027-632-7907	Sequence 7907, Ap
c 21	23	0.5	668	16	US-10-027-632-7907	Sequence 7907, Ap
22	23	0.5	678	13	US-10-424-599-102431	Sequence 102431,
23	23	0.5	976	9	US-09-764-877-3761	Sequence 3761, Ap
24	23	0.5	976	16	US-10-242-515-3761	Sequence 3761, Ap
25	23	0.5	1004	13	US-10-424-599-126393	Sequence 126393,
26	23	0.5	1734	17	US-10-437-963-29229	Sequence 29229, A
27	23	0.5	1782	13	US-10-424-599-137972	Sequence 137972,
28	23	0.5	1913	13	US-10-424-599-30424	Sequence 30424, A
29	23	0.5	2532	16	US-10-104-047-326	Sequence 326, App
30	23	0.5	4634	11	US-09-244-805-58	Sequence 58, Appl
31	23	0.5	4634	11	US-09-245-277-58	Sequence 58, Appl
32	23	0.5	4654	13	US-10-221-714A-508	Sequence 508, App
33	23	0.5	4654	15	US-10-311-455-2196	Sequence 2196, Ap
34	23	0.5	6311	15	US-10-311-455-1936	Sequence 1936, Ap
c 35	23	0.5	133632	13	US-10-087-192-1810	Sequence 1810, Ap
c 36	23	0.5	151858	17	US-10-322-281-653	Sequence 653, App
37	23	0.5	203654	9	US-09-820-905-3	Sequence 3, Appli
38	23	0.5	227246	17	US-10-322-281-314	Sequence 314, App
39	22	0.5	182	9	US-09-867-701-30	Sequence 30, Appl
40	22	0.5	256	9	US-09-867-701-3003	Sequence 3003, Ap
41	22	0.5	294	10	US-09-918-995-3607	Sequence 3607, Ap
42	22	0.5	338	13	US-10-424-599-139473	Sequence 139473,
43	22	0.5	362	15	US-10-198-846-12626	Sequence 12626, A
44	22	0.5	391	15	US-10-058-053A-175	Sequence 175, App
c 45	22	0.5	396	9	US-09-825-294-139	Sequence 139, App

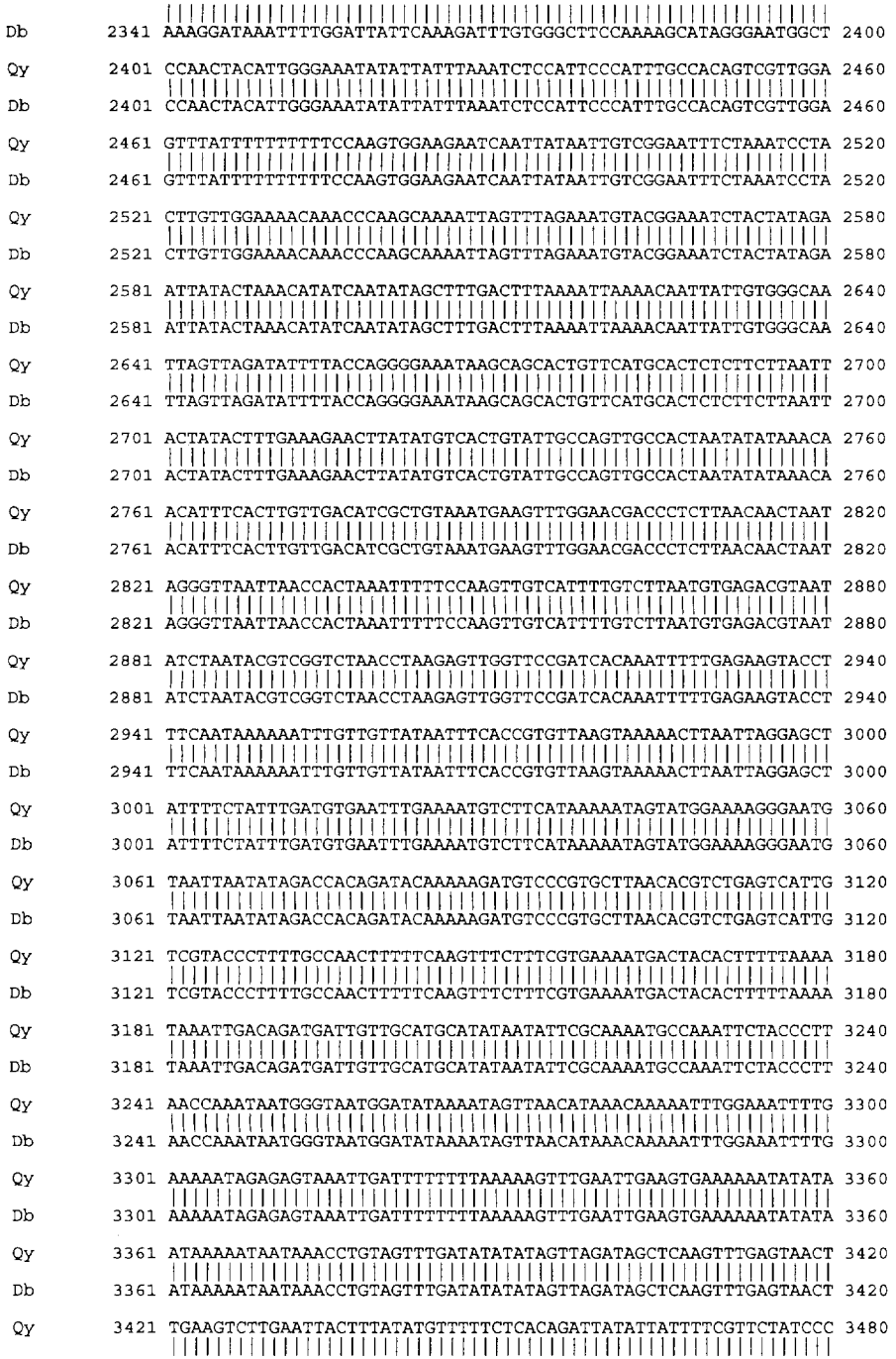
ALIGNMENTS

RESULT 1
US-10-027-880-3
; Sequence 3, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
; FILE REFERENCE: CHEP:003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4413
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-027-880-3

Query Match 100.0%; Score 4413; DB 15; Length 4413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGATCCACAGTGAATAAATAATAAGAACGGATTCGGTGATATTGCAACTATATAATGAA 60

Qy         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTGAATACTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATACTCTA 120

Qy        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAATAAATATCTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGG 180
```


Qy 181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240
Db 181 ATATATAAACAACCGGCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTT 240
Qy 241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300
Db 241 TAGGATTTACTTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTAT 300
Qy 301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360
Db 301 CAAAAAAAAAGTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGC 360
Qy 361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420
Db 361 TAATTTAAAGCCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATAT 420
Qy 421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480
Db 421 GCACCCATTTGTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAA 480
Qy 481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540
Db 481 CAATATCCATAAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATA 540
Qy 541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600
Db 541 TACAATACCACGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTT 600
Qy 601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660
Db 601 TGGTTTTATTTTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTA 660
Qy 661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720
Db 661 AATTTCGAATGTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCA 720
Qy 721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780
Db 721 CAGGGACATGCATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAA 780
Qy 781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 840
Db 781 AGGAAACTTCCTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCA 840
Qy 841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900
Db 841 GAAGAAAACGTATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCG 900
Qy 901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960
Db 901 AGACGCTGACTGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGG 960
Qy 961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020
Db 961 GTCAAAACCATATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGC 1020
Qy 1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080
Db 1021 TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTA 1080
Qy 1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140
Db 1081 CCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGC 1140
Qy 1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200
Db 1141 CCAAAATTTGAAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGT 1200
Qy 1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260
Db 1201 TAAATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAAC 1260
Qy 1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320
Db 1261 ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATT 1320
Qy 1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380
Db 1321 GTTGTATGGATTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGT 1380
Qy 1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440
Db 1381 TTATATAGCAGAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGC 1440
Qy 1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500
Db 1441 GTAGTCCTTTCTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTG 1500
Qy 1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560
Db 1501 TGTAATTCATAGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTC 1560
Qy 1561 TTTGCAACAACAAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620
Db 1561 TTTGCAACAACAAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTT 1620
Qy 1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680
Db 1621 GTTGTCCTGTCGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTC 1680
Qy 1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740
Db 1681 CAAGATTATAATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTAT 1740
Qy 1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800
Db 1741 CCATATAATTTCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCAC 1800
Qy 1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860
Db 1801 CCCCTTGCTAATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGT 1860
Qy 1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920
Db 1861 TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAA 1920
Qy 1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980
Db 1921 ATTCAACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGT 1980
Qy 1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040
Db 1981 AAATATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGAT 2040
Qy 2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100
Db 2041 ATAAATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTC 2100
Qy 2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160
Db 2101 GTTTTGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATC 2160
Qy 2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
Db 2161 AAACATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
Qy 2221 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280
Db 2221 GATGACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAA 2280
Qy 2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340
Db 2281 TCAGATTAATTTGTTTAATACGATAATGTATTTTGATTAAGTAGCTCTCAGCCAACAGGC 2340
Qy 2341 AAAGGATAAATTTTGGATTATTCAAAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400



RESULT 2
US-10-027-880-4
; Sequence 4, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE

```
;  TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
;  TITLE OF INVENTION:  VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
;  TITLE OF INVENTION:  PROMOTER AND TRANSGENIC PLANTS OBTAINED
;  FILE REFERENCE: CHEP:003US
;  CURRENT APPLICATION NUMBER: US/10/027,880
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: PCT/FR00/01768
;  PRIOR FILING DATE: 2000-06-23
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 4309
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: pBin19 vector
;   OTHER INFORMATION: insert
US-10-027-880-4

  Query Match            48.8%;  Score 2154;  DB 15;  Length 4309;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2154;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        131 CTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAAC 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 CTGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAAC 84

Qy        191 AACCGGCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTAC 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 AACCGGCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTAC 144

Qy        251 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 TTTGTAATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAA 204

Qy        311 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 GTACAAGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAG 264

Qy        371 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 CCTAATCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTT 324

Qy        431 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GTTTGTCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCAT 384

Qy        491 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 AAGAAGTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCA 444

Qy        551 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CGGTCTCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATT 504

Qy        611 TTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 TTATTCTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAAT 564

Qy        671 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 GTGCCTAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATG 624

Qy        731 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 CATAGGGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTC 684

Qy        791 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACG 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 CTAAAAGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACG 744

Qy        851 TATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        745 TATGGCCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 804

Qy        911 TGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCA 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TGATTTTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCA 864

Qy        971 TATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTA 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 TATCCAATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTA 924

Qy       1031 TTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTC 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 TTCGTGATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTC 984

Qy       1091 TACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTG 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 TACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTG 1044

Qy       1151 AAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGA 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 AAGAGTTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGA 1104

Qy       1211 ATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACA 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 ATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACA 1164

Qy       1271 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1165 TTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGA 1224

Qy       1331 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1225 TTAAATTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCA 1284

Qy       1391 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTT 1450
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1285 GAAAATATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTT 1344

Qy       1451 CTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCAT 1510
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1345 CTCCTCATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCAT 1404

Qy       1511 AGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAA 1570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1405 AGTCTTTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAA 1464

Qy       1571 CAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGT 1630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1465 CAAAAAAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGT 1524

Qy       1631 CGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATA 1690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1525 CGAAATCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATA 1584

Qy       1691 ATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 ATCTTTTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1644

Qy       1751 TCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTA 1810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1645 TCTAACATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTA 1704

Qy       1811 ATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAA 1870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1705 ATGTTTTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAA 1764

Qy       1871 TGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAA 1930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1765 TGGTCCAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAA 1824

Qy       1931 CATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTG 1990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1825 CATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTG 1884
```

```
Qy       1991 TTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACA 2050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1885 TTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACA 1944

Qy       2051 TTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAA 2110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1945 TTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAA 2004

Qy       2111 ATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATG 2170
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2005 ATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATG 2064

Qy       2171 ATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2230
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2065 ATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2124

Qy       2231 GTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2125 GTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2178


RESULT 3
US-10-027-880-1
; Sequence 1, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
;  APPLICANT: HOFFMAN, BEATE
;  TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
;  TITLE OF INVENTION:  VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
;  TITLE OF INVENTION:  PROMOTER AND TRANSGENIC PLANTS OBTAINED
;  FILE REFERENCE: CHEP:003US
;  CURRENT APPLICATION NUMBER: US/10/027,880
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: PCT/FR00/01768
;  PRIOR FILING DATE: 2000-06-23
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2149
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-10-027-880-1

  Query Match             48.7%;  Score 2149;  DB 15;  Length 2149;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy        196 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy        556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy        616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy        676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy        736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy        796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy        856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy        916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy        976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy       1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy       1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy       1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy       1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy       1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy       1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy       1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy       1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy       1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy       1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy       1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy       1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy       1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy       1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy       1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy       1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

RESULT 4
US-10-027-880-2
; Sequence 2, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
; FILE REFERENCE: CHEP:003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4280
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Construction
; OTHER INFORMATION: promoter + sequence coding for the gus gene
US-10-027-880-2

Query Match 48.7%; Score 2149; DB 15; Length 4280;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        136 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACAACCG 60

Qy        196 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGAAAGTACAAGTTCTACCTTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT 120

Qy        256 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTCCTGAATCTTTCATTTCTTGAATTGATATTTACATTTTTATCAAAAAAAAAGTACA 180

Qy        316 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTTCTACCAAAGCACAGGAGTTAAACAACTTGTGTGTCAAATGCTAATTTAAAGCCTAA 240

Qy        376 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTATGATTTCCCTTTTCTTCACGATATATACTGATATTGATATGCACCCATTTGTTTG 300

Qy        436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360

Qy        496 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTATATTGTGAAAAAAAAAAAAAAAAAGTGGTATACTGGTATATACAATACCACGGTC 420

Qy        556 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCGAATTGCCTCAACAATTTCTAGGAGAAAATGGACGTGTCTCTTTGGTTTTATTTTATT 480

Qy        616 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTTAATAACATACTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTCGAATGTGCC 540

Qy        676 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTCGAAGCCACAGGGACATGCATAG 600

Qy        736 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCACGTAGTTACCTTTAAAACCATCAAAAATATATTAATAGAAAAGGAAACTTCCTAAA 660

Qy        796 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAACAATTTAATAAAGTGGATAAAAAAAGATAAGAAGGTAGGCAGAAGAAAACGTATGG 720

Qy        856 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGCGACTCGTAACAAGGGACGTCCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780

Qy        916 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTTCTTTTTCTTTTCCTAAAGAACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC 840

Qy        976 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTGTTCTGCCTATTATTATATAACTAAAGATCCCCTCTTGTGCTTTGTCTTTATTCGT 900

Qy       1036 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATATATAATCTAACTTAAATTAGTTCTAAAATATATATGTCCTACCTATGTTTCTACTG 960

Qy       1096 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACGCCCAAAATTTGAAGAG 1020

Qy       1156 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1021 TTCCTCTTCCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080

Qy       1216 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAAACATGATAGACATTTAG 1140

Qy       1276 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTAAAACTTGAAATGTTATTTGAACTCTTTGGATTACGTGGATTGTTGTATGGATTAAA 1200

Qy       1336 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTGAAGATATTTATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAA 1260

Qy       1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTTGCGTAGTCCTTTCTCCT 1320

Qy       1456 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CATCCTCCTATGAAGAAAAATCCAAATAGTTTAAGGAAATTTTTGTGTAATTCATAGTCT 1380

Qy       1516 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTTTCGTAACCACAGTTCTATGTAGCTATCGTCATCATATTCCTCTTTGCAACAACAAAA 1440

Qy       1576 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGATCGTTTTTGTAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500

Qy       1636 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCTAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCCAAGATTATAATCTT 1560

Qy       1696 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

Qy       1756 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CATATACGTCTTGTTTTTGGTAGGCTCTGCGTCTTTTGAGACCACCCCCTTGCTAATGTT 1680

Qy       1816 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGTTGCACCTTAGACAATCCATAATACGTTACGTGAGTCGAAGTTGCACCAAAATGGTC 1740

Qy       1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTCAACAAACATGC 1800

Qy       1936 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAATATTCTGTTTAA 1860

Qy       1996 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 2055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTA 1920

Qy       2056 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980

Qy       2116 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAACATGATGATTAA 2040

Qy       2176 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGGGTCAT 2100

Qy       2236 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2284
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTAATCAG 2149
```

```
RESULT 5
US-10-311-455-2073/c
; Sequence 2073, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2073
;   LENGTH: 15046
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2073

  Query Match             0.6%;  Score 26;  DB 15;  Length 15046;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3349 AAAAAATATATAATAAAAATAATAAA 3374
              ||||||||||||||||||||||||||
Db       9895 AAAAAATATATAATAAAAATAATAAA 9870


RESULT 6
US-10-087-192-1681
; Sequence 1681, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1681
;   LENGTH: 68140
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(68140)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1681

  Query Match             0.6%;  Score 25;  DB 13;  Length 68140;
  Best Local Similarity   100.0%;  Pred. No. 3.8;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        507 TGAAAAAAAAAAAAAAAAAGTGGTAT 531
              |||||||||||||||||||||||||
```

```
Db      47578 TGAAAAAAAAAAAAAAAAGTGGTAT 47602


RESULT 7
US-09-918-995-8408
; Sequence 8408, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8408
;   LENGTH: 388
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-8408

  Query Match             0.5%;  Score 24;  DB 10;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 TATTGTGAAAAAAAAAAAAAAAAG 525
              ||||||||||||||||||||||||
Db        260 TATTGTGAAAAAAAAAAAAAAAAG 283


RESULT 8
US-10-424-599-58606/c
; Sequence 58606, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 58606
;   LENGTH: 440
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_23933C.1
US-10-424-599-58606

  Query Match             0.5%;  Score 24;  DB 13;  Length 440;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        501 ATATTGTGAAAAAAAAAAAAAAAA 524
              ||||||||||||||||||||||||
Db         75 ATATTGTGAAAAAAAAAAAAAAAA 52


RESULT 9
US-10-310-154-361
; Sequence 361, Application US/10310154
; Publication No. US20030233670A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Chomet, Paul S.
;  APPLICANT:  Adams, Thomas H
;  APPLICANT:  Ruff, Thomas G.
;  APPLICANT:  Agarwal, Ameeta K.
;  APPLICANT:  Ahrens, Jeffrey E.
;  APPLICANT:  Ball, James A.
;  APPLICANT:  Banu, G.
;  APPLICANT:  Bell, Erin
;  APPLICANT:  Boddupalli, Raghava
;  APPLICANT:  Deikman, Jill
;  APPLICANT:  Deng, Molian
;  APPLICANT:  Dong, Jinzhuo
;  APPLICANT:  Duff, Stephen M.
;  APPLICANT:  Galligan, Meghan M.
;  APPLICANT:  Hinchey, Brenda S.
;  APPLICANT:  Huang, Shihshieh
;  APPLICANT:  Johnson, G. Richard
;  APPLICANT:  Jung, Vincent
;  APPLICANT:  Kretzmer, Keith A
;  APPLICANT:  Laccetti, Lucille B.
;  APPLICANT:  Lai, Chao-Qiang
;  APPLICANT:  Lee, Gary
;  APPLICANT:  Lin, Jie-Yi
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  Lu, Bin
;  APPLICANT:  Luethy, Michael M.
;  APPLICANT:  Lund, Adrian
;  APPLICANT:  Madson, Linda L.
;  APPLICANT:  Malloy, Kathleen A.
;  APPLICANT:  McKiel, Christine L.
;  APPLICANT:  Miller, Philip W.
;  APPLICANT:  Padmavathi, Manchikanti
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Start, William G.
;  APPLICANT:  Tennesen, Dan
;  APPLICANT:  Vidya, K.R.
;  APPLICANT:  Wang, Haiyun
;  APPLICANT:  Xin, Zhanguo
;  APPLICANT:  Xu, Nanfei
;  APPLICANT:  Yang, Chunzhi
;  APPLICANT:  Zeng, Xiaoping
;  APPLICANT:  Zhang, Qiang
;  APPLICANT:  Zhao, Yajuan
;  APPLICANT:  Zhou, Li
;  TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
;  FILE REFERENCE: 38-15(52796)B
;  CURRENT APPLICATION NUMBER: US/10/310,154
;  CURRENT FILING DATE:  2002-12-04
;  PRIOR APPLICATION NUMBER: 60/337,358
;  PRIOR FILING DATE: 2001-12-04
;  NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 361
;   LENGTH: 2211
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (93)..(1634)
;   OTHER INFORMATION:
US-10-310-154-361

  Query Match             0.5%;  Score 24;  DB 16;  Length 2211;
  Best Local Similarity   100.0%;  Pred. No. 8.1;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 TATTGTGAAAAAAAAAAAAAAAAG 525
              ||||||||||||||||||||||||
Db       2188 TATTGTGAAAAAAAAAAAAAAAAG 2211


RESULT 10
US-10-221-613-238
; Sequence 238, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 238
;   LENGTH: 8964
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-238

  Query Match             0.5%;  Score 24;  DB 13;  Length 8964;
  Best Local Similarity   100.0%;  Pred. No. 9;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3350 AAAAATATATAATAAAAATAATAA 3373
             ||||||||||||||||||||||||
Db       7648 AAAAATATATAATAAAAATAATAA 7671


RESULT 11
US-10-311-455-1508
; Sequence 1508, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1508
;   LENGTH: 8964
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1508

  Query Match             0.5%;  Score 24;  DB 15;  Length 8964;
  Best Local Similarity   100.0%;  Pred. No. 9;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3350 AAAAATATATAATAAAAATAATAA 3373
             ||||||||||||||||||||||||
Db       7648 AAAAATATATAATAAAAATAATAA 7671


RESULT 12
US-10-087-192-1132/c
; Sequence 1132, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1132
;   LENGTH: 126266
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(126266)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1132

  Query Match             0.5%;  Score 24;  DB 13;  Length 126266;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        509 AAAAAAAAAAAAAAAAAGTGGTATA 532
              ||||||||||||||||||||||||
Db      59913 AAAAAAAAAAAAAAAAAGTGGTATA 59890


RESULT 13
US-09-954-556-17/c
; Sequence 17, Application US/09954556
; Publication No. US20030078219A1
; GENERAL INFORMATION:
;  APPLICANT: Brett P. Monia
;  APPLICANT:  Susan M. Freier
;  APPLICANT:  Scott Cooper
;  TITLE OF INVENTION: ANTISENSE MODULATION OF FIBROBLAST GROWTH FACTOR RECEPTOR 2 EXPR
;  FILE REFERENCE: RTS-0250
;  CURRENT APPLICATION NUMBER: US/09/954,556
;  CURRENT FILING DATE:  2001-09-14
;  NUMBER OF SEQ ID NOS: 108
; SEQ ID NO 17
;   LENGTH: 132762
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(132762)
;   OTHER INFORMATION: n = A,T,C or G
US-09-954-556-17

  Query Match             0.5%;  Score 24;  DB 10;  Length 132762;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        509 AAAAAAAAAAAAAAAAAGTGGTATA 532
              ||||||||||||||||||||||||
Db      54339 AAAAAAAAAAAAAAAAAGTGGTATA 54316


RESULT 14
US-09-803-719-1086/c
; Sequence 1086, Application US/09803719
```

```
; Publication No. US20030044783A1
; GENERAL INFORMATION:
;  APPLICANT: Williams, Lewis T.
;  APPLICANT:  Escobedo, Jaime
;  APPLICANT:  Innis, Michael A.
;  APPLICANT:  Garcia, Pablo Dominiguez
;  APPLICANT:  Sudduth-Klinger, Julie
;  APPLICANT:  Reinhard, Christoph
;  APPLICANT:  Giese, Klaus
;  APPLICANT:  Randazzo, Filippo
;  APPLICANT:  Kennedy, Giulia C.
;  APPLICANT:  Pot, David
;  APPLICANT:  Kassam, Altaf
;  APPLICANT:  Lamson, George
;  APPLICANT:  Drmanac, Radoje
;  APPLICANT:  Crkvenjakov, Radomir
;  APPLICANT:  Dickson, Mark
;  APPLICANT:  Drmanac, Snezana
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Leshkowitz, Dena
;  APPLICANT:  Kita, David
;  APPLICANT:  Garcia, Veronica
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Stache-Crain, Birgit
;  TITLE OF INVENTION: Human Genes and Gene Products
;  FILE REFERENCE: 1624.002
;  CURRENT APPLICATION NUMBER: US/09/803,719
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: 60/188,609
;  PRIOR FILING DATE: 2000-03-09
;  NUMBER OF SEQ ID NOS: 2396
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1086
;   LENGTH: 326
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(326)
;   OTHER INFORMATION: n = A,T,C or G
US-09-803-719-1086

  Query Match             0.5%;  Score 23;  DB 10;  Length 326;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        509 AAAAAAAAAAAAAAAAAGTGGTAT 531
              |||||||||||||||||||||||
Db        280 AAAAAAAAAAAAAAAAAGTGGTAT 258


RESULT 15
US-10-114-774-13
; Sequence 13, Application US/10114774
; Publication No. US20030166533A1
; GENERAL INFORMATION:
;  APPLICANT: Wolfner, Mariana
;  APPLICANT:  Lung, Oliver
;  APPLICANT:  Tram, Khanh-Uyen
;  TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
;  TITLE OF INVENTION:  MELANOGASTER
;  FILE REFERENCE: 19603/1791
;  CURRENT APPLICATION NUMBER: US/10/114,774
;  CURRENT FILING DATE:  2002-04-02
;  PRIOR APPLICATION NUMBER: 09/219,983
;  PRIOR FILING DATE: 1998-12-23
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 367
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-10-114-774-13

  Query Match             0.5%;  Score 23;  DB 15;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 TATTGTGAAAAAAAAAAAAAAA 524
              |||||||||||||||||||||||
Db        336 TATTGTGAAAAAAAAAAAAAAA 358


Search completed: July 31, 2004, 23:18:50
Job time : 1825 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 10:22:14 ; Search time 10330 Seconds
(without alignments)
12757.193 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatccacagtgaataaat.....tttcatgggttacgagtcca 4413

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	355	8.0	355	29	CC793681	CC793681 SALK_0173
	2	201	4.6	452	28	BH611124	BH611124 SALK_0299
	3	103	2.3	501	9	AV442527	AV442527 AV442527
	4	94	2.1	619	9	AV832242	AV832242 AV832242
c	5	75	1.7	379	28	BH757746	BH757746 SALK_0081
	6	73	1.7	575	9	AV552576	AV552576 AV552576
	7	67	1.5	74	28	BZ595332	BZ595332 SALK_0868
c	8	55	1.2	92	29	BX285678	BX285678 Arabidops
c	9	44	1.0	83	29	AL763344	AL763344 Arabidops
	10	30	0.7	703	28	BZ470713	BZ470713 BONNA88TR
c	11	30	0.7	833	28	BZ457686	BZ457686 BONRW80TR
c	12	29	0.7	874	28	BZ508310	BZ508310 BONHJ80TR
c	13	26	0.6	646	28	BH963752	BH963752 odf98b01.
c	14	26	0.6	691	28	BH942846	BH942846 ode19f10.
	15	26	0.6	736	28	BH438688	BH438688 BOGGM25TR
	16	26	0.6	1209	28	CC258283	CC258283 CH261-51M
c	17	25	0.6	128	14	CD420389	CD420389 ri86b01.y
c	18	25	0.6	281	9	AW016688	AW016688 UI-H-BI0p
c	19	25	0.6	300	14	CD606616	CD606616 RZ153A3E0
c	20	25	0.6	307	14	CD605249	CD605249 RK020A1D0
	21	25	0.6	329	9	AW004306	AW004306 701555389
c	22	25	0.6	350	9	AI838645	AI838645 UI-M-AO0-
c	23	25	0.6	388	14	CD420251	CD420251 ri84c02.y
c	24	25	0.6	410	10	AW575585	AW575585 UI-HF-BL0
	25	25	0.6	418	14	CF273396	CF273396 EST3085 Z
c	26	25	0.6	442	9	AA925170	AA925170 UI-R-A1-e
c	27	25	0.6	445	10	AW165189	AW165189 fe05e02.x
c	28	25	0.6	465	12	BM673424	BM673424 UI-E-CR1-
	29	25	0.6	478	13	BU764603	BU764603 sas03h09.
c	30	25	0.6	479	10	AW121179	AW121179 UI-M-BH2.
	31	25	0.6	579	14	CB851263	CB851263 UI-CF-DU1
	32	25	0.6	592	10	AW424360	AW424360 sh64f05.y
	33	25	0.6	598	14	CD374151	CD374151 TNWbmfC11
c	34	25	0.6	603	12	BQ003582	BQ003582 UI-H-EI1-
c	35	25	0.6	619	13	BU238032	BU238032 Ds01_14c1
	36	25	0.6	637	14	CB588141	CB588141 AGENCOURT
c	37	25	0.6	639	14	CB823573	CB823573 EST 4797
c	38	25	0.6	681	13	BU628781	BU628781 UI-H-FG0-
c	39	25	0.6	685	13	BQ705470	BQ705470 em_Contig
	40	25	0.6	721	29	CE625066	CE625066 tigr-gss-
c	41	25	0.6	741	12	BQ008193	BQ008193 UI-H-ED1-
c	42	25	0.6	758	13	BU633081	BU633081 UI-H-FL1-
	43	25	0.6	2441	11	BC052222	BC052222 Danio rer
	44	24	0.5	96	9	AA432834	AA432834 vf28h11.r
c	45	24	0.5	142	12	BI133442	BI133442 UI-M-BH3-

ALIGNMENTS

RESULT 1
CC793681
LOCUS CC793681 355 bp DNA linear GSS 01-JUL-2003
DEFINITION SALK_017399.55.00.x Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_017399.55.00.x, genomic survey sequence.
ACCESSION CC793681
VERSION CC793681.1 GI:32388904
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 355)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

```
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES              Location/Qualifiers
     source           1. .355
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /strain="Columbia 0"
                      /db_xref="taxon:3702"
                      /clone="SALK_017399.55.00.x"
                      /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                      /note="PCR was performed on Arabidopsis thaliana lines
                      each of which contains  one or more TDNA insertion
                      elements. The resultant fragment for each  line was
                      directly sequenced to determine the genomic sequence at
                      the  site of insertion. Details of the protocols used can
                      be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             8.0%;  Score 355;  DB 29;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-76;
  Matches  355;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1925 AACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAAT 1984
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db             1 AACAAACATGCATCGTTTCAAATTTTATTTATTCAATGGCGTTATTTGTTCATTGTAAAT 60

Qy       1985 ATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAA 2044
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 ATTCTGTTTAACTCACTGACGAATTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAA 120

Qy       2045 ATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTT 2104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 ATAACATTTTATGGAACCACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCGTTT 180

Qy       2105 TGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAAC 2164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 TGTGAAATCATTAACGACCTACATTTGATCCCTCATTACTTTAATAATTAGGAATCAAAC 240

Qy       2165 ATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATG 2224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           241 ATGATGATTAAGTTCACCAAAGACGTCTCTTATGGCTATTAAGAGTCAGACGCAAGGATG 300

Qy       2225 ACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTA 2279
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           301 ACCGGGGTCATTAAGACGTCTTATATTCAACCATTACTCCACTAATTGCTAATTA 355


RESULT 2
BH611124
LOCUS       BH611124                452 bp    DNA     linear   GSS 04-JAN-2002
DEFINITION  SALK_029914 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_029914, genomic survey sequence.
ACCESSION   BH611124
VERSION     BH611124.1  GI:18058246
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within an annotated intron of At1g77750.
            Class: TDNA tagged.
FEATURES              Location/Qualifiers
     source           1. .452
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /strain="Columbia 0"
                      /db_xref="taxon:3702"
                      /clone="SALK_029914"
                      /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                      /note="PCR was performed on Arabidopsis thaliana lines
                      each of which contains  one or more TDNA insertion
                      elements. The resultant fragment for each  line was
                      directly sequenced to determine the genomic sequence at
                      the  site of insertion. Details of the protocols used can
                      be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             4.6%;  Score 201;  DB 28;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-39;
  Matches  201;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4188 ATTCTTAAGGAAGCATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTT 4247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           252 ATTCTTAAGGAAGCATCTCAAATGTTGGCTAGTTCTTGTCTCAGGTTAAAACAATGTTTT 311

Qy       4248 GCAATTTGCTTTTAGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTG 4307
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           312 GCAATTTGCTTTTAGTTAGAATTGTTGACTTGCTTAGCTTTTTGACTAACTCTGTCTCTG 371

Qy       4308 TGAAGCAAAGTTTGATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATC 4367
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           372 TGAAGCAAAGTTTGATCAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTATACATC 431

Qy       4368 AATTTATGTATTTTCTTTGTC 4388
             |||||||||||||||||||||
Db           432 AATTTATGTATTTTCTTTGTC 452


RESULT 3
AV442527
LOCUS       AV442527                501 bp    mRNA    linear   EST 14-NOV-2000
DEFINITION  AV442527 Arabidopsis thaliana above-ground organ two to six-week
            old Arabidopsis thaliana cDNA clone APZ12a05_r 5', mRNA sequence.
ACCESSION   AV442527
VERSION     AV442527.1  GI:7612939
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7 (3), 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES              Location/Qualifiers
     source           1. .501
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
```

```
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="APZ12a05_r"
                     /tissue_type="aboveground organs"
                     /dev_stage="two to six-week old"
                     /clone_lib="Arabidopsis thaliana above-ground organ two to
                     six-week old"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             2.3%;  Score 103;  DB 9;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-15;
  Matches  103;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3704 TTCACCGTCAAATCTCTGAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGAC 3763
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCACCGTCAAATCTCTGAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGAC 60

Qy       3764 GACCTTGTTCGACATCAGCCAGTCTCTGCTTCGTAATGTTACG 3806
             ||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTTGTTCGACATCAGCCAGTCTCTGCTTCGTAATGTTACG 103


RESULT 4
AV832242
LOCUS       AV832242                 619 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV832242 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-07-C12 5',
            mRNA sequence.
ACCESSION   AV832242
VERSION     AV832242.1  GI:19874302
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .619
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL11-07-C12"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL11"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to various
                     treatments (dehydration, cold, high salt, ABA, heat and
                     UV). Dark-grown plants"
ORIGIN
```

```
  Query Match             2.1%;  Score 94;  DB 9;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-13;
  Matches   94;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3713 AAATCTCTGAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTT 3772
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AAATCTCTGAAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTT 62

Qy       3773 CGACATCAGCCAGTCTCTGCTTCGTAATGTTACG 3806
             ||||||||||||||||||||||||||||||||||
Db         63 CGACATCAGCCAGTCTCTGCTTCGTAATGTTACG 96


RESULT 5
BH757746/c
LOCUS       BH757746                 379 bp    DNA     linear   GSS 01-MAR-2002
DEFINITION  SALK_008127.24.55.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_008127.24.55.x, genomic
            survey sequence.
ACCESSION   BH757746
VERSION     BH757746.1  GI:19042094
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_008127.24.55.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains one or more TDNA insertion
                     elements. The resultant fragment for each line was
                     directly sequenced to determine the genomic sequence at
                     the site of insertion. Details of the protocols used can
                     be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             1.7%;  Score 75;  DB 28;  Length 379;
  Best Local Similarity   100.0%;  Pred. No. 1e-08;
  Matches   75;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        132 TGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACA 191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 TGCAGTCGAATTGTGATATATTGTAAGCAATCTGAAAAGAATAAGTGGGATATATAAACA 158

Qy        192 ACCGGCGAAAGTACA 206
              |||||||||||||||
Db        157 ACCGGCGAAAGTACA 143
```

```
RESULT 6
AV552576
LOCUS       AV552576                575 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  AV552576 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ33e03R 5', mRNA sequence.
ACCESSION   AV552576
VERSION     AV552576.1  GI:8723989
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7 (3), 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZ33e03R"
                     /tissue_type="roots"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             1.7%;  Score 73;  DB 9;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-08;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3734 GACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCAGCCAGTCTCTGCT 3793
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCAGCCAGTCTCTGCT 60

Qy       3794 TCGTAATGTTACG 3806
             |||||||||||||
Db         61 TCGTAATGTTACG 73


RESULT 7
BZ595332
LOCUS       BZ595332                 74 bp    DNA     linear   GSS 07-JAN-2003
DEFINITION  SALK_086892.43.40.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_086892.43.40.x, genomic
            survey sequence.
ACCESSION   BZ595332
VERSION     BZ595332.1  GI:27535851
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 74)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA. This sequence lies within an annotated intron of At1g77750.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .74
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_086892.43.40.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

  Query Match             1.5%;  Score 67;  DB 28;  Length 74;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-06;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4347 TCTCTCTTTAGATTATACATCAATTTATGTATTTTCTTTGTCTACAGTTTCATGGGTTAC 4406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCTCTCTTTAGATTATACATCAATTTATGTATTTTCTTTGTCTACAGTTTCATGGGTTAC 60

Qy       4407 GAGTCCA 4413
             |||||||
Db         61 GAGTCCA 67


RESULT 8
BX285678/c
LOCUS       BX285678                 92 bp    DNA     linear   GSS 07-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-384F07-017271,
            genomic survey sequence.
ACCESSION   BX285678
VERSION     BX285678.1  GI:28884674
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 92)
  AUTHORS   Rosso,M., Strizhov,N., Li,Y. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2003) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion within the locus defined by clone T32E8. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
```

```
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .92
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-384F07-017271"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
ORIGIN

  Query Match             1.2%;  Score 55;  DB 29;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 0.0028;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2008 TTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACC 2062
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         92 TTTTTTAATTTTTCAAAGAAGAACATTTTTGATATAAATAACATTTTATGGAACC 38


RESULT 9
AL763344/c
LOCUS       AL763344                  83 bp    DNA     linear   GSS 19-JUN-2002
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-036B01-015001,
            genomic survey sequence.
ACCESSION   AL763344
VERSION     AL763344.1  GI:21512449
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 83)
  AUTHORS   Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion within the locus defined by clone T32E8. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .83
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-036B01-015001"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
ORIGIN

  Query Match             1.0%;  Score 44;  DB 29;  Length 83;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1638 TAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCC 1681
                ||||||||||||||||||||||||||||||||||||||||||||
Db         83 TAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCC 40


RESULT 10
BZ470713
LOCUS       BZ470713                  703 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONNA88TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONNA88,
            genomic survey sequence.
ACCESSION   BZ470713
VERSION     BZ470713.1  GI:26768062
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BONNA88TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONNA88"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             0.7%;  Score 30;  DB 28;  Length 703;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        882 GACCACTGCGGAGACGGCGAGACGCTGACT 911
                ||||||||||||||||||||||||||||||
Db        660 GACCACTGCGGAGACGGCGAGACGCTGACT 689


RESULT 11
BZ457686/c
```

```
LOCUS       BZ457686                 833 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONRW80TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONRW80,
            genomic survey sequence.
ACCESSION   BZ457686
VERSION     BZ457686.1  GI:26737311
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BONRW80TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .833
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONRW80"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             0.7%;  Score 30;  DB 28;  Length 833;
  Best Local Similarity   100.0%;  Pred. No. 3.5e+02;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        882 GACCACTGCGGAGACGGCGAGACGCTGACT 911
              ||||||||||||||||||||||||||||||
Db        647 GACCACTGCGGAGACGGCGAGACGCTGACT 618


RESULT 12
BZ508310/c
LOCUS       BZ508310                 874 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  BONHJ80TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONHJ80,
            genomic survey sequence.
ACCESSION   BZ508310
VERSION     BZ508310.1  GI:27031022
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BONHJ80TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONHJ80"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             0.7%;  Score 29;  DB 28;  Length 874;
  Best Local Similarity   100.0%;  Pred. No. 5.9e+02;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2314 TGATTAAGTAGCTCTCAGCCAACAGGCAA 2342
              |||||||||||||||||||||||||||||
Db        301 TGATTAAGTAGCTCTCAGCCAACAGGCAA 273


RESULT 13
BH963752/c
LOCUS       BH963752                 646 bp    DNA     linear   GSS 01-OCT-2002
DEFINITION  odf98b01.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH963752
VERSION     BH963752.1  GI:23444978
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: odf98  row: b  column: 01
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 64
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
ORIGIN

  Query Match             0.6%;  Score 26;  DB 28;  Length 646;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+03;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        870 AAGGGACGTCCCGACCACTGCGGAGA 895
              ||||||||||||||||||||||||||
Db        515 AAGGGACGTCCCGACCACTGCGGAGA 490


RESULT 14
```

```
BH942846/c
LOCUS       BH942846                691 bp    DNA     linear   GSS 01-OCT-2002
DEFINITION  ode19f10.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH942846
VERSION     BH942846.1  GI:23422906
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: ode19  row: f  column: 10
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
ORIGIN

  Query Match             0.6%;  Score 26;  DB 28;  Length 691;
  Best Local Similarity   100.0%;  Pred. No. 3.9e+03;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        283 TGATATTTACATTTTTATCAAAAAAA 308
              |||||||||||||||||||||||||
Db        610 TGATATTTACATTTTTATCAAAAAAA 585


RESULT 15
BH438688
LOCUS       BH438688                736 bp    DNA     linear   GSS 12-DEC-2001
DEFINITION  BOGGM25TR BOGG Brassica oleracea genomic clone BOGGM25, genomic
            survey sequence.
ACCESSION   BH438688
VERSION     BH438688.1  GI:17624402
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGGM25TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGGM25"
                     /clone_lib="BOGG"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             0.6%;  Score 26;  DB 28;  Length 736;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+03;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        283 TGATATTTACATTTTTATCAAAAAAA 308
              |||||||||||||||||||||||||
Db        192 TGATATTTACATTTTTATCAAAAAAA 217


Search completed: July 31, 2004, 18:43:15
Job time : 10338 secs
```